

Historical, Legal and Political Issues

Year XX, No. 2-3 / 2024.

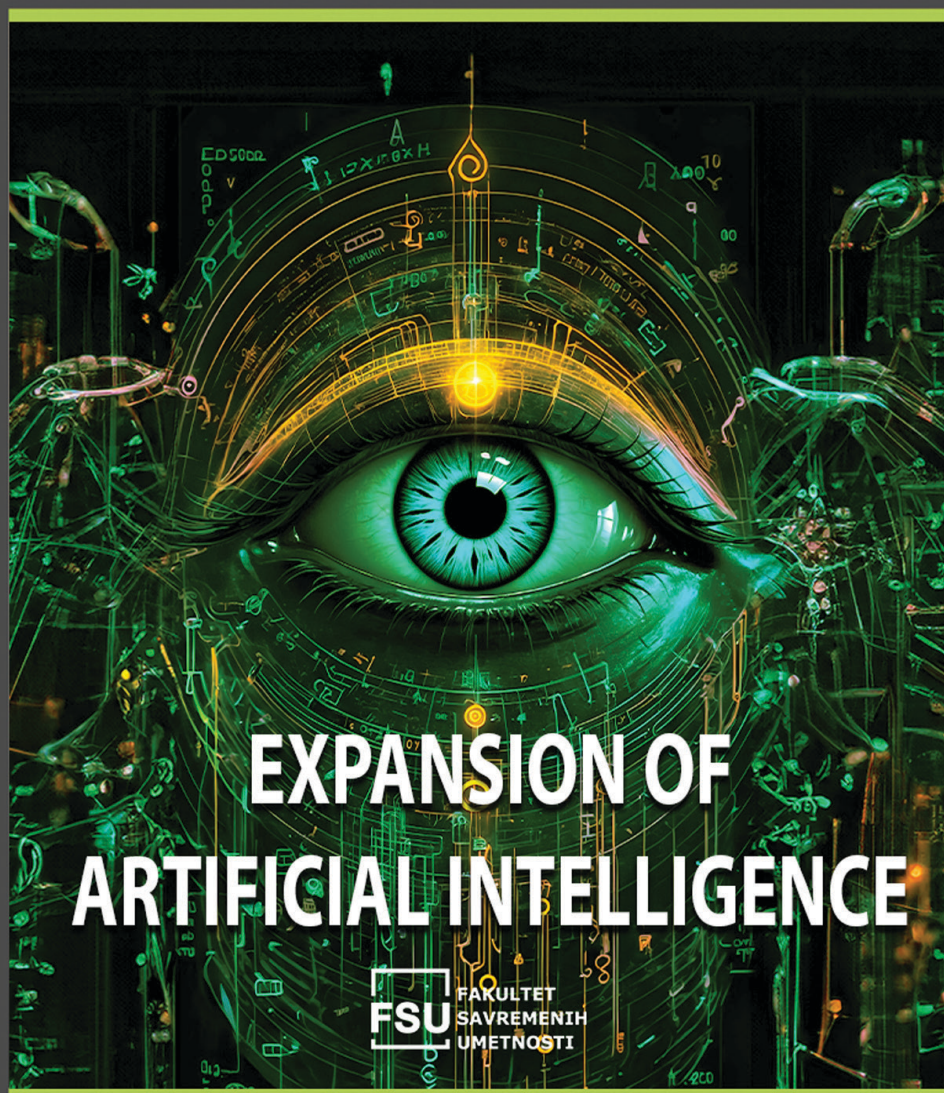
LIMES

LIMES
PLUS

JOURNAL OF SOCIAL SCIENCES AND HUMANITIES

ISSN 1820-0869

EXPANSION OF ARTIFICIAL INTELLIGENCE



LIMES

PLUS

JOURNAL OF SOCIAL SCIENCES AND HUMANITIES

EXPANSION OF ARTIFICIAL INTELLIGENCE

Issue editor
Viktor Radun

FSU FAKULTET
SAVREMENIH
UMETNOSTI

ISSN 1820-0869

Issue 2–3/2024

Scientific Editorial Board:

Ljiljana DOBROVŠAK, Ivo GOLDSTEIN, Eric GORDY, Egidio IVETIĆ, Dušan JANJIĆ, Isidora JARIĆ, Anđelka MIHAJLOV, Aleksandar MIRKOVIĆ, Marc PILKINGTON, Ognjen RADONJIĆ, Agnieszka SADECKA, Christine SINAPI, Pritam SINGH, Aleksandra STUPAR, Darko TANASKOVIĆ, Faruk ULGEN, Josip VRANDEČIĆ

Publisher:

HERAedu, Belgrade
Faculty of Philosophy



Faculty of Contemporary Arts
Belgrade, Svetozara Miletića 12
journal@limes-plus.com
<https://limes-plus.com/>
<https://www.fsu.edu.rs/>

For the Publisher:

*Zorica Stablović BULAJIĆ
Prof. dr Dragan ČALOVIĆ*

Editor-in-Chief:

Nikola SAMARDŽIĆ

Executive Editorial Board:

*prof. dr Valentin KULETO
prof. dr Milena ILIĆ (sekretar)
prof. dr Dragan ČALOVIĆ
Zorica Stalović BULAJIĆ*

Proofreading and Copyediting (Serbian):

LINKgroup Belgrade

Proofreading and Copyediting (English):

LINKgroup Belgrade

Technical Editing:

LINKgroup Belgrade

Cover Photo:

Ana ŽIVKOVIĆ

Subscription:

*journal@limes-plus.com;
+381605522581*

Printing:

Instant system, Belgrade

AUTHORS:

Zorana BODIROGA

*Faculty of Political Sciences,
University of Belgrade, Serbia*

Jelena BOŠKOVIĆ

*Metropolitan University,
Belgrade, Republic of Serbia*

Sladana ČABRILO

*Department of International
Business Administration, I-Shou
University, Kaohsiung, Taiwan*

Stevica DEĐANSKI

*Faculty of Social Sciences, FNP,
Belgrade, Serbia*

Aleksandar S. ĐORĐEVIĆ

*Institute of Comparative Law,
Terazije 41, 11000 Belgrade,
Serbia*

Dragana Vujović ĐERMANOVIĆ

*General Director of BeeShaper,
Serbia*

Aleksandar FILIPOVIĆ

*University of Montenegro,
Faculty of Medicine, Podgorica,
Montenegro*

Boris JEVTIĆ

*School of Computing, RAF,
Union University Belgrade,
Serbia*

Svetlana KRALJ

*Independent Researcher,
Podgorica, Montenegro*

Tatjana Mamula NIKOLIĆ

*Metropolitan University,
Belgrade, Serbia*

Sanja MARKOVIĆ

*Academy of Vocational Studies
of Kosovo and Metohija,
Department of Zvečan, Serbia*

Filip MARKOVIĆ

*Faculty of Technical
Sciences, Kosovska
Mitrovica, Serbia*

Lam V. NGUYEN

*Department of International
Business Administration,
I-Shou University, Kaohsiung,
Taiwan*

Marko POPOVIĆ

*"Svetozar Marković"
Gymnasium, Novi Sad,
Republic of Serbia*

Vera POPOVIĆ

*Institute of Field and Vegetable
Crops, Novi Sad, Republic of
Serbia*

Semir POTURAK

*Faculty of Technical Studies in
Travnik, University of Travnik,
Bosnia and Herzegovina*

Viktor RADUN

*Faculty of Contemporary Arts,
Belgrade, University of Business
Academy, Novi Sad, Serbia*

Dejan RANČIĆ

*Faculty of Electronic
Engineering, Niš, Serbia*

Aleksandar ŠEVIĆ

*TCD, Trinity Business School,
Dublin, Ireland, College of
Vocational Studies for IT,
Belgrade, Serbia*

Željko ŠEVIĆ

*International University of
Applied Sciences, Erfurt,
Thuringia, Germany*

Milan SIMIĆ

RMIT University, Australia

Žaklina SPALEVIĆ

*Singidunum University Belgrade,
Serbia*

Ivan TRIFUNOVIĆ

*Tourism Organization of
Vrnjačka Banja, Vrnjačka Banja*

Hanh T. My VU

*Postgraduate Management
Program, I-Shou University,
Kaohsiung, Taiwan*

CIP - Каталогизacija у публикацији
Народна библиотека Србије, Београд
32

LIMES plus : journal of Social Sciences and Humanities
/ editor-in-chief Nikola Samardžić. - [Štampano izd.]. -
2004, br. 1- . - Beograd : HERAedu, 2004- (Beograd
: Instant system). - 24 cm

Tri puta godišnje. - Drugo izdanje na drugom mediju:
Limes plus (Online) = ISSN 2406-2871
ISSN 1820-0869 = Limes plus
COBISS.SR-ID 114047756

TABLE OF CONTENTS

FOREWORD

VIKTOR RADUN: QUO VADIS, VI?

Part I: Dimensions and Challenges in Creating Intelligent Systems

9 Slađana ČABRILO, Hanh T. MY VU, Lam V. NGUYEN, Dragana VUJOVIĆ ĐERMANOVIĆ:
AI in content marketing: advantages and disadvantages in the AI era

25 Aleksandar ŠEVIĆ, Željko ŠEVIĆ:
AI in the financial industry: how to prevent data breaches

43 Vera POPOVIĆ, Jelena BOŠKOVIĆ, Aleksandar FILIPOVIĆ i Marko POPOVIĆ:
Application of artificial intelligence (AI) in biotechnology and medicine

73 Semir POTURAK,
Integration of artificial intelligence in professional architectural practice

Part II: Ethical, Legal, and Social Aspects of Artificial Intelligence

99 Viktor RADUN:
Ethical implications and social challenges of artificial intelligence development towards artificial general intelligence

121 Nataša ĐURĐEVIĆ, Marija MAKSIMOVIĆ i Tatjana MAMULA NIKOLIĆ:
Irresponsible use of ChatGPT impedes academic performance

149 Aleksandar S. ĐORĐEVIĆ, Stevica DEĐANSKI, Boris JEVTIĆ:
Challenges of human rights protection and security in the age of artificial intelligence

171 Zorana BODIROGA:
Application of artificial intelligence in detection and prevention of peer violence: a literature review

Part III:
Artificial Intelligence at Your Service

193 Stevica DEDANSKI, Boris JEVTIĆ:

The impact of AI-supported services on customer experience: a case study of the hotel industry in Serbia

215 Svetlana KRALJ, Tatjana MAMULA NIKOLIĆ:

Digital technologies, AI and consumer behavior: the influence of influencers

235 Ivan TRIFUNOVIĆ, Žaklina SPALEVIĆ i Sanja MARKOVIĆ:

Progress of artificial intelligence: challenges and opportunities in e-tourism

253 Ivan TRIFUNOVIĆ, Žaklina SPALEVIĆ, Dejan RANČIĆ, Filip MARKOVIĆ, Milan SIMIĆ:

Application of artificial intelligence in detecting fraud in tourism

EXPANSION OF ARTIFICIAL INTELLIGENCE

Artificial Intelligence – AI, originally conceived as a scientific and technological field within computer science with the goal of producing intelligent machines capable of performing operations that humans can, has traveled a long way in a short amount of time. Although the concept of AI originated as far back as the 1950s, it remained relatively unknown outside of specialized circles until recently. Numerous technological advancements over the past decades, particularly in recent years of the 21st century, have contributed to the fascinating progress toward achieving programs whose intelligence could not only match but also surpass average human intelligence. From being almost entirely unknown just a few years ago, the field of artificial intelligence has exploded to become a significant social phenomenon and an inevitable mainstream topic discussed not only in scientific and professional circles but also among the general public.

Today, artificial intelligence imposes itself as a daily reality, something omnipresent yet invisible and intangible. It is becoming an independent factor radically reshaping not only the economy, business practices, and other areas of society but also ways of thinking, human culture, education, creativity, and even the very nature of humanity and the essence of human existence, with unimaginable consequences for the vision of society and humanity in the future.

Given the relevance of this topic and, on the other hand, the necessity of understanding the incredible expansion of artificial intelligence into our established everyday lives, which radically changes our perception of society and reality as a whole, this issue of the journal presents works by prominent authors who examine both theoretical and practical challenges and questions related to the functioning, development, and application of AI in various aspects and areas of the economy and society.

This issue of the journal *Limes+*, with the theme “The Expansion of Artificial Intelligence,” includes a total of twelve papers by authors from Serbia, Bosnia and Herzegovina, Montenegro, Australia, Ireland, Germany, and Taiwan, including three papers by authors from other countries and nine papers from Serbia. Based on the thematic closeness of the topics they address, all the papers are grouped into three sections: Part I – Dimensions and Challenges in Creating Intelligent Systems; Part II – Ethical, Legal, and Social Aspects of Artificial Intelligence; and Part III – Artificial Intelligence at Service.

In the first paper within the first section of this issue, which aims to explore various dimensions and challenges of creating AI systems in different domains of the economy and society, Slađana Čabrilo, Hanh T. My Vu, Lam V. Nguyen, and Dragana Vujović Đermanović discuss the advantages and weaknesses of applying artificial intelligence in content marketing. In the paper authored by Vera Popović, Jelena Bošković, Aleksandar Filipović, and Marko Popović, the potentials of AI are summarized and their applications in medicine, agriculture, and biology-based industries are analyzed. The paper by Aleksandar and Željko Šević examines the challenges faced by financial institutions in the AI era concerning data theft, cyber-attacks, and fraudulent activities enabled by the misuse of artificial intelligence, offering possible solutions and responses. Finally, in the concluding paper of this section, Semir Poturak analyzes the application of artificial intelligence in professional architectural practice and construction. Poturak examines the integration of AI and information and communication technologies into the process of architectural design, exploring possibilities for the implementation of new technologies in these fields.

In the second section of the issue, dedicated to the ethical, legal, and social aspects of artificial intelligence, Viktor Radun investigates the ethical implications and societal challenges of AI development, focusing on the transition to artificial general intelligence and the need for its social control. Nataša Đurđević, Marija Maksimović, and Tatjana Mamula Nikolić examine the impact of AI tools like ChatGPT on the overall learning process, with a particular focus on the irresponsible use of ChatGPT, including changes and consequences for education. In the paper by Aleksandar S. Đorđević, Stevica Dedanski, and Boris Jevtić, the challenges posed by artificial intelligence are explored, aiming to improve the quality of work, collaboration, and experiences of all participants in the process of protecting human rights and security supported by AI technologies. Lastly, Zorana Bodiroga's paper deals with the possibilities of applying artificial intelligence in the detection and prevention of peer violence.

In the third section of this issue, titled "Artificial Intelligence at Service," papers explore the application of artificial intelligence in the service sector. In their paper, Stevica Dedanski and Boris Jevtić, through exact research, point out how AI-supported services in hospitality influence improving the user experience. Svetlana Kralj and Tatjana Mamula Nikolić conducted a study aiming to analyze the mutual influence of digital technologies, artificial intelligence, influencers, and consumer behavior. Finally, two papers address the use of artificial intelligence in tourism. In the first, Žaklina Spalević, Ivan Trifunović, and Sanja Marković explore the development of the artificial intelligence market and the role of legal regulations in AI applications in tourism, analyzing the challenges and opportunities of implementing AI in this field. In the second paper, Žaklina Spalević, Ivan Trifunović, Dejan Rančić, Filip Marković, and Milan Simić examine the role of artificial intelligence, particularly machine learning and deep learning, in fraud detection and ensuring secure financial transactions in e-tourism.

PART I

DIMENSIONS AND CHALLENGES IN CREATING INTELLIGENT SYSTEMS

AI IN CONTENT MARKETING: ADVANTAGES AND DISADVANTAGES IN THE AI ERA

Sladjana Cabrilo¹

*Department of International Business Administration, I-Shou
University, Kaohsiung, Taiwan*

Hanh T. My Vu²

*Postgraduate Program in Management, I-Shou University, Kaohsiung,
Taiwan*

Lam V. Nguyen³

*Department of International Business Administration, I-Shou
University, Kaohsiung, Taiwan*

Dragana Vujovic Djermanovic⁴

CEO of BeeShaper, Serbia

Abstract: *Artificial intelligence (AI) is rapidly changing many industries, including content marketing. AI tools can automate various content creation and distribution tasks, allowing marketers and content creators to increase*

1 sladjana@isu.edu.tw

2 vivianhanh279@gmail.com

3 ilfer27032001@gmail.com

4 gagadjermanovic@gmail.com

their efforts. However, it also brings new threats and challenges that need to be addressed. This project examines the key ways AI influences content marketing strategies. More specifically, it aims to understand the benefits and drawbacks of AI in content marketing and anticipate the future of AI integration into content marketing. The research collected data through a structured interview conducted with a marketing expert and then compared gathered insights regarding the advantages and disadvantages of AI in content marketing with ChatGPT-generated responses. The main findings reveal that efficiency and imagination enhancement are the main advantages of AI in content marketing, with the lack of human touch, data privacy and ethical concerns as the disadvantages. The study contributes to the current theoretical discussions on AI in content marketing, providing insights that are beneficial for marketers and helping marketing managers refine strategies for using AI more effectively in content marketing.

Keywords: *AI tools, content marketing, advantages, disadvantages, interview, ChatGPT.*

INTRODUCTION

In today's competitive business world, people want to finish their tasks quickly. Computers are used a lot, which creates a ton of data. Artificial Intelligence (AI) refers to the ability of computer systems to perform tasks that typically require human intelligence, such as recognizing visual images, understanding speech, making decisions, and translating languages (Chintalapati and Pandey, 2022). In the marketing industry, Artificial Intelligence Marketing (AIM) is an approach of optimally utilizing technology and customer data to enhance the customer's experience (Chintalapati and Pandey, 2022). Content marketing is a marketing strategy that focuses on sharing useful information to

attract and keep the attention of a specific group of people (Jain and Aggarwal, 2020). Content marketing aims to change or improve customer behavior by consistently providing high-quality content through AIM as an assistant. In marketing, creating and sharing content is very important. The rise of digital platforms and the attention economy has made it crucial to have engaging content that helps build brand awareness and communicates value. However, it can be hard for brands to handle the large amount of text and media data. It is important to use AI responsibly to get the most out of it and avoid problems. More research is needed to understand both the benefits and risks of using AI in content marketing as it continues to grow and change.

The main research question of this research is: What are the advantages and disadvantages of the implementation of AI in content marketing in the AI era?

In order to answer the main research question, we use the following sub-questions:

Q1: How is AI used in content marketing?

Q2: What are the differences between human-based and AI-based insights in content marketing?

LITERATURE REVIEW

Changes in marketing in the AI era

In today's dynamic digital marketing environment, businesses encounter numerous challenges, including managing vast datasets, adapting to changing consumer preferences and competing intensely. The advent of AI is fundamentally transforming marketing strategies, empowering personalized customer interactions, streamlining data analysis processes, and facilitating automated decision-making. This transition towards AI-driven marketing signifies a significant shift towards utilizing data-driven insights to inform decision-making processes.

Integration of AI in Content Marketing

Content marketing is a marketing strategy that involves the creation and distribution of online content such as videos, blog articles, social media posts, and other materials which are used to attract, and engage an audience and customers. (Lintulahti, 2023). AI technologies are revolutionizing content creation by providing marketers with powerful tools to streamline the process, enhance efficiency, and improve the quality of content. Nowadays, AI has totally changed how we do content marketing. Think of AI as a really smart helper who makes creating, sharing, and understanding our content much easier. First, AI can quickly create different types of content using something called Natural Language Generation. This means we can make articles, blogs, or even product descriptions super-fast. And it is not just any content – it has content that is made specifically for each person based on what they like and do online. AI looks at a bunch of data from different places to figure out what people are interested in. Once we have our content, AI helps us share it in the best way possible. It knows the right time and place to post stuff so more people see it. And if someone has a question about our content, AI-powered chatbots can step in and help them, making it seem like we are available all the time. Moreover, AI helps us find cool things from the internet to share with our audience, which keeps them engaged and interested. In short, AI has made content marketing much simpler and more effective, and it is only getting better with time (Chintalapati and Pandey, 2022).

Advantages of AI in Content Marketing

AI in content marketing can be very useful because it permits the generation of ideas and the automatic creation of text about almost everything. For example, marketers can ask Chat GPT to create a short text about content to inspire them. It is also possible to change the nature of the text and for example make it more

funny, professional, or more understandable for kids. AI bridges communication gaps. It simplifies complex jargon, making content accessible to wider audiences. Nowadays AI can also generate videos, images, music, design, and text. The creation possibilities are very wide. AI can also analyze many databases such as user behavior, preferences, and demographics to segment audiences effectively. It tailors content to specific groups, enhancing engagement. For instance, AI can create personalized product recommendations based on browsing history or send targeted email campaigns with relevant content. AI is also a great tool for repetitive and annoying tasks. AI handles mundane chores, such as scheduling social media posts, and keyword optimization. It liberates marketers from routine tasks, allowing them to focus on strategic planning. AI in marketing is a real assistant that can assist in every daily task for short-term and long-term missions and make inconvenient tasks more convenient (Lintulahti, 2023).

Disadvantages of AI in Content Marketing

Lack of human touch and authenticity

It is important to remember, though, that everything ChatGPT writes or creates is based on what it understands about something that has been written before. This means it is not actually capable of original thought or creativity in the same way as humans (Hauser, Li, and Mao, 2023).

AI has trouble being truly original and creative like humans because it does not have human experiences, deep knowledge, or understanding. It mostly just combines existing information from the internet. AI cannot set goals, imagine, or improve ideas like humans can over time. So, while AI can copy some creative tasks, humans are still better at real innovation. Humans have skills like imagination that AI cannot replicate. We can understand things in ways that do not make sense to AI. When AI does creative tasks, they often lack a human touch and seem strange or unsettling. Imagine you are trying to write a story. You can draw from your own life experiences and feelings to make it

interesting and unique. But AI does not have those experiences to draw from, so it relies on information it finds on the internet. This means its stories might feel a bit stale or unoriginal like they are just copying what is already out there.

Privacy And Ethical Concerns

AI in content marketing uses your personal info to tailor advertisements and suggestions, which can be a privacy problem. It looks at what you do online, buy, and share on social media to target you better. But this can feel like an invasion of privacy because it has been done without asking you first. Also, if the AI learns from biased data, it might unfairly target certain groups. Since there are no rules on how companies can use your information, there is worry about manipulation. To fix this, marketers need clear rules on how AI can use personal data and ways to make sure it is fair for everyone.

Create a strategy to integrate AI into Content Marketing

Creating a strategy to integrate AI into Content Marketing is very essential in the AI era. It can make AI more reliable for marketers while using AI tools for their work. Here are three stages of a circle of content improvement that marketers or the content team can use to improve effective AI integration into content marketing.

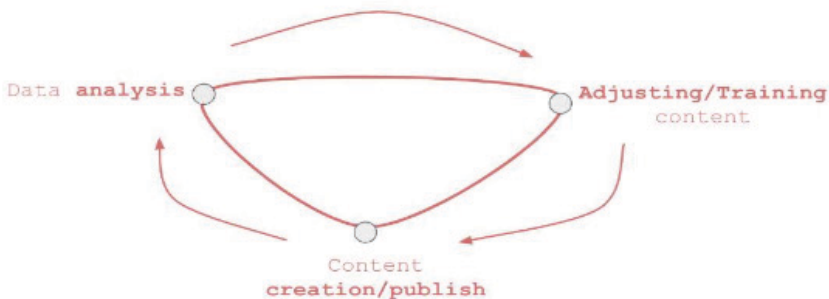


Figure 1. Circle of Content Improvement (Lintulahti, 2023)

First Stage: Content Creation

The content team first looks at what they already have and thinks up new ideas that match their marketing goals. Then, they figure out who we are trying to reach, study what they want, and come up with more ideas, even looking at what our competitors are doing. After that, they make the content using special AI tools and check it to make sure it fits our brand, what our customers like, and what's popular right now.

Second Stage: Gathering and Analyzing Data

After publishing content, the team keeps an eye on how it is doing across different platforms. They look at things like engagement rate, and bounce rate to figure out how well it is performing. Then, they dive into the data to see what trends and insights they can find. This helps them decide what kind of content to make in the future, based on what works best for the audience.

Third Stage: Adjustment and Training

Data analysis guides content adjustments, getting in on successful content types, and optimizing traffic sources for effectiveness. The team undergoes specialized training to enhance their AI content creation skills and catch up with current trends and technologies in the field.

METHOD

Data collection

We first conducted a structured interview with an esteemed expert in marketing, with rich experience spanning over a decade. The expert has been actively contributing to the field of marketing. With a strong foundation in both academia and practical experience, he has been a great source of valuable insights regarding our main research questions. We tried to approach the expert when we were conducting the project about content marketing in March 2024, and by sending an email with a list of questions to the expert we already prepared before about content marketing, we got the answers from the experts that are presented in Table 1.

ChatGPT is a virtual assistant or chatbot that uses a field of machine learning known as natural language processing (NLP). It is an example of generative AI because it can create something completely new that has never existed before. It can engage in conversations, answer questions, generate text, and assist with various tasks. Trained on a diverse dataset from the internet, ChatGPT can understand and produce human-like responses across different subjects. Its applications range from customer service to content creation and education. However, ethical considerations regarding misinformation and bias are important in its use. Overall, ChatGPT represents a significant advancement in AI-driven communication and interaction (Hauser, Li, and Mao, 2023).

The reason we chose ChatGPT to answer the list of questions about AI in Content Marketing is that ChatGPT nowadays is one of the most famous and approachable AI tools all over the world. Almost everyone knows about ChatGPT and uses it for their work, studies, and research. Although precise details of the training data have not been made public, it's known that it includes a massive dataset of 175 billion parameters, including books, newspapers, and research papers, and a snapshot of the internet taken in 2021. It was trained using unsupervised learning, meaning that the information it was given was not labeled. Instead, it simply records which words are used and analyzes the context and relationships between them. This allows it to then predict what words, and in what order, it should use to create the best response. Although the results it is capable of may seem amazing, the algorithms that make up ChatGPT are really quite simple and are all based on statistics. The "intelligence" is all possible due to the sheer size of the training dataset and the speed at which it is capable of processing requests (Hauser, Li, and Mao, 2023). That is the reason why we chose ChatGPT to answer the list of questions about Content Marketing we prepared before.

Results

All results are presented in Table 1.

Questions	Expert's answer	ChatGPT's answer
1. What role does AI play in content marketing in general?	"AI significantly transforms content marketing by streamlining content creation, personalization, and analysis. It aids marketers in generating tailored content, predicting consumer behavior, and automating routine tasks, thereby allowing them to concentrate more on creative and strategic aspects of marketing."	AI plays a significant role in various aspects of content marketing, including content creation, optimization, personalization, distribution, and analysis. It enables marketers to understand audience preferences, create targeted content, automate repetitive tasks, and improve overall efficiency.
2. What are the benefits of using AI in content marketing?	"The benefits of integrating AI into content marketing include heightened efficiency through automation, enhanced personalization capabilities leading to better engagement, valuable insights gained from data analysis to inform strategies, improved search engine optimization, and content performance optimization through real-time testing."	Scalable Content: AI produces content for diverse audiences. Personalization: AI enhances engagement with tailored content. Efficiency: AI automates tasks, saving time and resources. Insightful Analytics: AI-driven analytics optimize content strategies.

Questions	Expert's answer	ChatGPT's answer
3. What are the drawbacks or challenges associated with AI in content marketing?	"Drawbacks of using AI in content marketing include the potential loss of the personal touch in content, concerns regarding data privacy and ethics, high initial implementation costs, the complexity of AI tools requiring a steep learning curve, and the risk of generating content that may not always be accurate or contextually appropriate without human oversight".	Quality Concerns: AI-generated content may lack authenticity. Ethical Issues: Concerns about data privacy and bias. Cost and Complexity: Implementation can be expensive. Data Dependency: AI relies on accurate and unbiased data.
4. How do you suggest addressing the disadvantages or challenges previously mentioned?	"To mitigate these challenges, it is recommended to maintain human oversight over AI-generated content, prioritize data privacy and compliance with regulations, evaluate the potential ROI before investing, invest in training for effective AI tool utilization, and adopt a hybrid approach that combines the strengths of AI with human creativity and insight."	Human Oversight: Combine AI automation with human creativity. Ethical Guidelines: Establish clear ethical guidelines. Training: Provide training for effective AI usage. Monitoring: Regularly monitor AI systems for biases and errors.
5. What are your predictions for the future of AI in content marketing?	"Predictions for the future of AI in content marketing include deeper integration across all aspects of content marketing, the emergence of more advanced personalization techniques, improvements in the quality of AI-generated content to closely resemble human writing, an increasing focus on ethical use and regulation of AI technologies, and the creative blending of AI with other emerging technologies to create new content marketing opportunities."	Enhanced Personalization: Deeper levels of content personalization. Sophisticated Content Creation: AI-generated content becomes more sophisticated. Ethical Focus: Increased attention to ethical AI usage. Integration: AI integrates with emerging technologies for immersive experiences.

Table 1. Answer from interviewing the expert and ChatGPT

Data Analysis

Data received from the expert and ChatGPT-generated was analyzed to identify similarities and differences that are presented in Chapter IV.

DISCUSSION

The expert and ChatGPT responses share many similarities but differ in clarity, specificity, and speed. The expert's answers are easy to understand, specific, and creative, while ChatGPT's responses are faster but sometimes use advanced language, requiring more time for comprehension and a lack of future prediction and human judgment.

Question 1: Role of AI in Content Marketing

Similarities: Both mention AI's role in content creation, personalization, and analysis, and its help in automating tasks and focusing on creative and strategic marketing aspects.

Differences: The expert says AI helps make content and predicts what people will like and uses it as an assistant tool to help marketing managers plan marketing strategies, while ChatGPT means that AI can help make content better, send it out, and figure out what people want. So, while the expert looks at the potential of AI, ChatGPT wants to cover more areas where AI helps in content marketing.

Question 2: Benefits of using AI in Content Marketing

Similarities: Both answers highlight efficiency through automation, enhanced personalization, and valuable insights from data analysis.

Differences: Expert adds improved search engine optimization and real-time testing for content performance, whereas ChatGPT mentions scalable content production and insightful analytics to optimize content strategies.

Question 3: Drawbacks/Challenges of AI in Content Marketing

Similarities: Both acknowledge the potential lack of human touch brain and data privacy concerns.

Differences: Expert discusses high initial costs and the complexity of AI tools, while ChatGPT points out quality concerns with AI-generated content (reality) and the dependency on accurate data.

Question 4: Addressing drawbacks/challenges

Similarities: Both suggest maintaining human oversight and prioritizing data privacy.

Differences: Expert recommends evaluating ROI before investing and adopting a hybrid approach, while ChatGPT advises establishing ethical guidelines, providing training, and monitoring AI systems for biases.

Question 5: Predictions for the future of AI in Content Marketing

Similarities: Both show a deeper integration of AI in content marketing and an increase in ethical use and regulation.

Differences: Expert predicts improvements in AI-generated content quality and creative blending with other technologies, while ChatGPT expects more sophisticated content creation and integration with emerging technologies for immersive experiences.

Based on previous studies, comparison results and Table 1, the main advantages of AI in content marketing are firstly efficiency because it helps

to generate new ideas, content, pictures, text, and more for humans. The role of a human worker is to use his brain to innovate or create new ideas then he can ask AI to generate them by writing, drawing, analyzing, and classifying among others. The human worker only needs to use his imagination which cannot be done by AI because AI comes from the human, so it does not have yet the possibility to do more than what he knows or what he is made for. So, the human is the brain and AI is the physical worker that quickly executes and achieves missions and repetitive tasks that allow the human to focus on the core of ideas, projects, innovations, or inventions. Secondly, AI is an advantage to help increase imagination and possibilities for humans. With AI we can quickly access most of the knowledge in the world, so because many things have already been done in the past, humans can use it as inspiration. It helps him to increase his imagination and become a better thinker.

Based on previous studies, comparison results and Table 1, AI in content marketing has some drawbacks. One big problem is that it is not as good as humans at being creative and reliable. Even though AI can make lots of content quickly, it often lacks the special understanding and originality that people have. Also, sometimes AI content feels too general and does not connect well with certain groups of people. Another worry is keeping people's data private and being ethical when using AI. People wonder if AI will change the jobs humans do in the future and, if so, what those jobs might be. Therefore, AI should be viewed as an assistant rather than a replacement for human roles. While AI might automate some tasks, there are still things humans are better at, like understanding feelings, being creative, and solving hard problems. So, even though AI has a lot of potential, we need to be careful about how we use it.

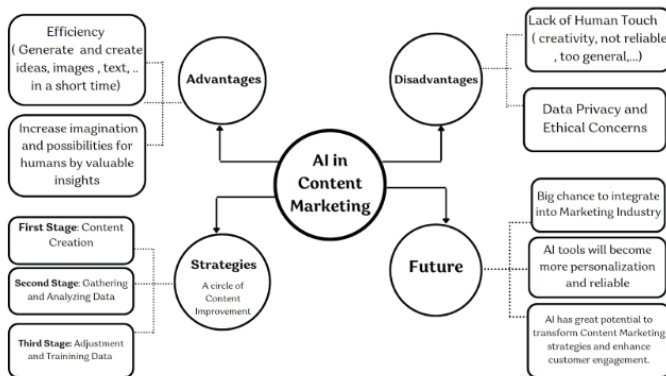


Figure 2. General findings of integration of AI in Content Marketing

CONCLUSION

In conclusion, this project has examined the advantages and disadvantages of using AI in content marketing. It is clear that AI brings significant benefits by automating routine tasks, generating personalized content at scale, and providing valuable insights through data analysis. These capabilities allow marketers to increase productivity and focus more on strategic and creative work. However, AI also faces limitations compared to human creativity, such as human creativity and reality. There are also ethical concerns regarding data privacy and potential biases.

The integration of artificial intelligence (AI) into content marketing has significant contributions to both theoretical and practical aspects of the marketing field. In marketing theoretical aspects, AI brings super-fast ideas, inspiring human workers with its ability to generate new concepts and offer unique viewpoints. By automating routine tasks like data analysis and content creation, AI acts as a helpful assistant, giving marketers more time to concentrate on important decisions and strategic planning. With AI's

support, marketers can get into higher-level thinking, ultimately leading to more innovative and effective marketing strategies. In practical marketing aspects, AI enhances content creation processes by analyzing vast amounts of data, identifying trends, and optimizing content. It ensures precision and relevance. AI also can create new content, including articles, social media posts, and product descriptions. It generates personalized content at scale. Overall, AI analyzes user behavior, preferences, and demographics to tailor marketing messages. It ensures that content reaches the right audience and right customers.

Throughout the project, we got some limitations that we realized, after finishing, could have been addressed better if we had the opportunity to start again. Firstly, we did not have enough time to interview multiple experts, which would have made our findings more reliable and clearer. Secondly, the list of questions we had prepared was not careful or deep enough for our topic and this project. Finally, lack of time and research experience significantly limit the quality of our findings.

Looking to the future, AI will become more integrated across the content marketing process. AI tools for content creation and personalization will become more reliable. If developed and applied carefully with appropriate guidelines, AI has great potential to transform content marketing strategies and enhance customer engagement. At the same time, human judgment remains crucial, especially for strategic planning, brand positioning, and evaluating qualitative aspects like tone and messaging. An optimal approach combines the strengths of AI for automation and analytics with human creativity, perspective, and decision-making. With balanced and responsible development, AI will serve to assist, not replace, the expertise of content marketing professionals. Continued research on its applications and impacts is still needed to maximize benefits and address new challenges.

REFERENCES

1. Adams, P. 2023. Coke Joins AI Hype Train In Search For Next-Gen Creative Capabilities. Marketing Dive. February 23, 2023. <https://www.marketingdive.com/news/coca-cola-OpenAI-Bain-ChatGPT-AI-marketing/643381/>.
2. AIContentfy. 2023. The Role Of Artificial Intelligence In The Future Of Content Marketing. February 8, 2023. <https://aicontentfy.com/en/blog/role-of-artificial-intelligence-in-future-of-content-marketing>.
3. Chintalapati, Srikrishna, and Shivendra Kumar Pandey. 2022. “Artificial Intelligence in Marketing: A Systematic Literature Review.” International Journal of Market Research 64 (1): 38–68. <https://doi.org/10.1177/14707853211018428>.
4. Gupta, Tarun, and Supriya Bansal. 2023. “Navigating the AI Revolution: A New Era in Marketing.” IJSR 12 (12): 2065–73. <https://doi.org/10.21275/sr24302221714>.
5. Hauser, John R., Zelin Li, and Chengfeng Mao. 2023. “Artificial Intelligence and User-Generated Data Are Transforming How Firms Come to Understand Customer Needs,” 147–67. <https://doi.org/10.1108/s1548-643520230000020007>.
6. Lintulahti, Niilo. 2023. “Creating a Strategy for AI Integration in Content Marketing.”
7. Norris, Phil. 2024. 26 Impressive Examples Of AI In Marketing - Social Media... Social Media Strategies Summit Blog. April 3, 2024. <https://blog.socialmediastrategiessummit.com/10-examples-of-ai-in-marketing>
8. Peterson, T. 2024. Transforming Content Marketing Lifecycles With AI | MarTech. MarTech. March 22, 2024. <https://martech.org/transforming-content-marketing-lifecycles-with-ai/>.
9. Statista. 2024. Marketing Related Use Of AI Worldwide By Region 2018 L Statistic | Statista. Accessed June 18, 2024. <https://www.statista.com/statistics/915372/marketing-related-use-artificial-intelligence-world/>

AI IN THE FINANCIAL INDUSTRY: HOW TO PREVENT DATA BREACHES

Aleksandar Šević¹

TCD, Trinity Business School, Dublin, Ireland

Information Technology School – ITS, Belgrade, Serbia

Željko Šević²

*International University of Applied Sciences, Erfurt, Thuringia, FR
of Germany*

Abstract: *The EU lags behind major US and Asian competitors in information security expenses per IT spending or employee. Compared to other areas of economic activities, financial institutions seem to be more complacent regarding potential system breaches. However, they are also better equipped to quickly identify inconsistencies because their infrastructure seems comparatively less complex. With the prospective application of NIS 2 Directives, it is imperative to examine the challenges financial institutions face with the current increase in AI capabilities that hackers and other malevolent actors could use without impunity if proper countermeasures and systems are not deployed. In this study, we provide an overview of challenges that financial institutions could face, but we also discuss solutions, such as advancements in generative AI products or the widespread application of behavioral biometrics, which should increase the reliability of online activities and prevent the abuse of personal data on a*

1 a.sevic@tcd.ie; aleksandar.sevic@its.edu.rs

2 Zeljko.Sevic@IU.org

broader scale. Apart from specific company solutions, we provide a discussion platform for government decision-makers.

Keywords: *artificial intelligence, banking, financial institutions, cybersecurity, behavioral biometrics.*

1. INTRODUCTION

By 17 October 2024, the EU member states are required to adopt measures necessary for the implementation of the NIS 2 Directive (EU 2022/2555). It replaces the NIS Directive from 2016 to ameliorate cybersecurity and establish reporting requirements across various sectors. In addition, the Digital Operation Resilience Act (DORA), adopted in September 2023, is sector-specific and takes precedence in some areas over the NIS Directive for financial entities (NIS 2, 2023).

Following these legal requirements, EU banks and other financial institutions should implement technical, operational and organizational measures to secure the network and information systems. While there may be occasional incursions by hackers, not only via bank servers but also via ATMs, it is also essential to underscore that the increase in online banking activities, which allowed for the closure of offices, decline in the workforce, and the reduction of the ATM network, also contributed to an increase in identity thefts and inappropriate money transfers.

While comprehensive antivirus protection is essential for financial institutions, it is also evident that the negligent approach to security by banking clients represents one of the stumbling blocks preventing the creation of a robust system. While the use of PINs, passwords and cards is still prevalent, there is an apparent shift among cybersecurity experts to the implementation of biometrics that are supposed to be resilient to purchase, robbery, duplication or counterfeiting (Marani et al., 2023). If there is malevolent access to passwords

or cards, the same could be replaced as many times as necessary, but the low entropy of fingerprints that are compromised by criminals, and more prone to be repeated in the general population, would force banks to scan other biometrics, such as face or irises, which are more person-specific.

With the advancement of AI, especially in natural language processing (NLP) or drawing, the proliferation of fake news or videos has become available to amateurs without any specialized knowledge in coding or design. This would imply that criminal or state-sponsored actors are trained and equipped to create impeccable announcements, web designs, photos or videos. If this is the case, banks and other financial institutions must move to the next step of behavioural biometrics that would recognize patterns using deep machine learning. However, since humans change behavioural patterns with age, the AI must be able to identify these shifts, as well.

In the continuation of the paper, we discuss the growing relevance of artificial intelligence, and then in Section 3, we discuss various types of AI-driven frauds. In Section 4, we shall discuss the improvements in the financial services in the near future and then finalize the study with concluding remarks.

2. AI MATTERS

Artificial intelligence can carry out human tasks, and it is achieved by collecting, interpreting and learning from externally provided data (Craft, 2011; Haenleing and Kaplan, 2019). Apart from the popular images and movies that captivate public imagination, be it positive or negative, regarding AI, we face a gradualist view that machines will help humankind replace rather dull and repetitive jobs with a shift towards group work with machines leading to the ultimate augmented intelligence (De Cremer and Kasparov, 2021). This postulate is based on the fact that machines are fast but lack emotional intelligence, intuition and cultural sensitivity. If AI could mimic these features, then human labor would be in peril of being completely replaced by machines. Should these concerns be voiced by leading physicists, such as Stephen Hawking or

businesspersons, such as Bill Gates, the threat is incomparably more resonating among the audience (Kumar and Choudhury, 2022; Rawlinson, 2015). The speed of AI development and adoption will determine the rate of human obsolescence and its replacement.

Even at the current level of AI development, where almost none of the more profound human-like features can be teased out of the machines, the advancements in language processing, image enhancement and voice recognition create opportunities and challenges for businesses and regulators. Financial institutions have already dabbled in using robo-advisors, but with the advancement in NLP, the ability to provide quality customer services has tremendously improved. Law departments would deploy self-learning neural networks to scan regulations not only in the host country but globally and secure manageable answers at a fraction of an earlier timeframe. Banks already evaluate customers more comprehensively with a vast speed improvement, while the search for AI-enabled company headquarters or other facilities would reduce insider frauds, but also, in exchange for some additional loss in privacy, employees should be able to enter the building, enjoy the working environment and leave the premises under constant surveillance, which, in turn, would reduce other, more visible, control points (Noonan, 2023). Finally, we may witness a decline in the hiring rates as human labor becomes less useful for employers.

Where did this drive to boost productivity originate? Firstly, the goal of reaping the benefits of newcomers into the market is enticing enough for banks to implement AI-driven changes. Secondly, the threat by FinTech companies has become a reality, and if conservative policies are implemented, banks will run the risk of becoming moribund giants of inefficiencies. Thirdly, the business world yearns to face years of significant productivity increases. Except for 1997–2004, the IT-driven growth has failed to re-appear. The advancement might have been limited to specific areas of human activity that did not convert into growth, or the measuring was incorrect (Brynjolfsson et al., 2019). The additional claim

by these authors that some technological advancements such as nuclear power or flying cars failed to become as omnipresent is a moot point because AI is a game changer, perhaps even the ultimate chapter of human development, because it complements, improves on and then it is forecasted to outshine humans in almost every aspect imaginable.

It is doubtful whether similar claims would have been made should ChatGPT in 2019 instead of 1.5 billion parameters have as much as 175 billion, the same as ChatGPT-3 only four years later. This improvement did not materialize without costs. The ChatGPT training had a strong impact on the environment, with 502 metric tons of carbon emitted. Adding ChatGPT to the Bing search engine increased the consumption per search by five times. So far, global data centers have used approximately one percent of global electricity annually, and with the recent breakthrough, the relative electricity consumption rate will have doubled by 2030 (Cho, 2023). An increase in parameters allows artificial neural networks to perform tasks more accurately and rapidly, but it also increases energy consumption. The reduction of this negative trend could be achieved by exacting less accuracy for tasks where high precision is not relevant or by making further inroads in the spiking neural network research where calculations could be up to 280 times less energy intensive since neurons are either active, when necessary, or remain inactive without energy consumption (Dora, 2024).

While the underlying processes of AI neural networks are known and heavily researched, the ramifications of AI on humanity are still unknown, implying that the regulatory framework is still to be implemented. The hype has not subsided; leading, mainly US, companies are jockeying for the top position in this realm, but it does not mean that many other OECD countries have not already outlined some general strategies for AI advancements. In 2017, China adopted “The Next AI Development Plan”, aiming to become the global innovation center by 2030. Soon after that, the EU in 2018 and the US in 2019 promoted the development or leadership in AI. The amount of committed investments in China is matched

by the ability of US companies to raise capital in highly liquid domestic markets, but the ringfenced outlays for AI in France, Germany or the UK, as one of the leading European countries, are comparatively smaller (Figure 1) (Sheppard et al., 2018). The strength of Europe originates in its talent (Figure 2). Suppose Sam Altman, the controversial CEO of OpenAI, is correct in estimating that investments in AI will range between several hundred billion US\$ and US\$ 7 trillion. In that case, the dominating players will be in the US and China, not in Europe (Cornish, 2024).

Figure 1. Investment in AI companies

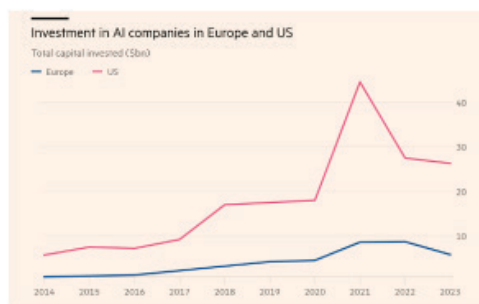


Figure 2. Number of workers hired



Source: Financial Times

On 13 March 2024, the European Parliament passed the first Artificial Intelligence Act in the world, aiming to protect safety, health and individual rights, but not to stem the growth in AI use in the EU. The Act prohibits using biometric categorization systems at work or in education to recognize emotional states or other vulnerabilities of subjects of interest. Social scoring or predictive policing is not allowed, as well. The areas to be explored in our paper will be placed under the aegis of high-risk AI systems that could be deployed in training, employment, private or public services. Under the new legislation, deploying detailed risk management, keeping logs, ensuring transparency and accuracy, and involving human oversight without exceptions are necessary.

Manipulating video, images or voice content must be clearly marked, and the more powerful general-purpose AI system will require more frequent and comprehensive reviews, coinciding with the level of damage it could inflict. The legislative act pays special attention to the certification process, taking into account technological progress, which means that at some point, high-risk AI systems may be evaluated as such (Artificial Intelligence Act, 2024). As a supranational organization, the EU legislation has to consider all member countries' interests, with specific exceptions, for instance, in Ireland, due to its relationship with the UK (Shmuha, 2021). It is a big step forward and should contribute to the preservation of fundamental rights that are enshrined in the previous EU legislation, but shortly after the adoption, Canada, Singapore and UAE are reported to have been approaching European AI start-ups and showering them with tax breaks, golden visas and other benefits (Criddle and Cornish, 2024). The UK has also joined the queue of bidders, but the closer the company is to the EU market, the less chance it has of being excluded from the EU Act if operations are conducted on the continent. In October 2023, the US President signed the Executive Order on "Safe, Secure, and Trustworthy" Artificial Intelligence, aiming to require various government agencies to prepare task lists ranging from 90 days to a year in the areas of safety and security, the use of AI by the US government, privacy protection, consumer protection and civil rights and equality (Rubin and Manes, 2024).

3. AI-DRIVEN FRAUDS

ChatGPT 3 from OpenAI, as the most conspicuous representative of a league of natural language processing products emanating from other competitors such as Microsoft Azure, Stanford CoreNLP, Google Cloud or IBM Watson, allowed many actors across the globe to maintain almost simultaneous "communication" with a machine. The 175 billion parameters allowed it to predict words at a speed rate and the context in which humans could assume that there is an intelligent counterpart, although somewhat constrained by engineers

and puzzling perceptions of ethics, on the other side. However, it was discovered very quickly that homework, essays and reports could be created within a very short period, thereby cutting the time necessary to complete the task. Then, the Dall-e 2, another product by OpenAI, offered enviable image generation. If, in such an abundance of neural network solutions, even amateurs at a fraction of the cost could manipulate photos, voices, videos and allegedly original documents, then the stage for the next-generation AI-driven fraud is unambiguously set.

The financial institutions and their customers will increasingly face threats such as:

- a) The proliferation of fabricated documents used in the process of identification or application submissions. The challenge is that these copies are so well made that internal checks and balances are insufficient to spot them.
- b) Fake voices over the phone or chat platforms could be used to impersonate close relatives, trick the person into providing payments, or mimic the voice of an abducted person in search of a ransom payment. If these are used to represent financial experts or managers involved in passing relevant decisions, a lack of trust, insecurity, or fraud could be perpetrated.
- c) Fake videos could be obtained from the dark web for US\$150 via WormGPT, FraudGPT, DarkBART or similar platforms. The deployment for phishing or ransomware is then further facilitated thanks to a multitude of products tailored for criminals. If videos impersonate celebrities, the scam becomes more convincing, and criminals can achieve targeted financial results with a relatively small sample of randomized victims. It may not be necessary to apply strategies to a broader audience, which may keep it outside the public's eye for longer.
- d) The easiest access to sensitive datasets is for company insiders. If artificial intelligence allows them to misuse company or client accounts with impunity, the incidence of breaches will increase.

- e) AI-powered algorithms in capital markets should be able to collude without the explicit violation of market watchdogs' rules. If price-trigger strategies are followed, algorithms switch from exploration to collusion with similar products, and at some point, significant profits will be recorded. Even in a high-noise environment where innovative trading strategies are difficult to follow, AI-supported algorithms demonstrate similar biases and again reap substantial profits. These processes may be already under way in capital markets and their final impact on markets and society is yet to be analyzed.
- f) Social engineering attacks account for four-fifths of all IT breaches and exploit human weaknesses by abusing psychological techniques. The AI-enhanced content via chatbots, voice generation, and fakes could make these attacks highly convincing and sophisticated, boosting the success rate.
- g) High-risk AI systems will become increasingly complex and brandish capabilities to access and manipulate personal data on a large scale. If these data uses cannot be ringfenced, there is a significant likelihood of leaking information to actors who should never be designated to have access (Venkataramakrishnan, 2024; Jaskiewicz, 2023; Dou et al., 2023; Zvelo, 2023; Shen, 2023).

4. FINANCIAL SERVICES OF THE NEAR FUTURE

Compared to transportation and aviation, communications, utilities and government, financial services have expressed the smallest increase in successful cyber attacks but also the second least concern about cyber warfare, trailing only behind communications. Despite this striking confidence, financial actors are confident that the majority of future attacks will originate in cloud services (identified by 46% of respondents), remote employees (39%) and insecure virtual private networks (VPN) (37%) (Bridewell, 2023). Since 2018, the cybersecurity risk, closely followed by the geopolitical risk, has become a major concern for

UK banks (Bank of England, 2023). These attitudes reflect the challenges that data-driven companies are about to face.

McKinsey outlines multiple steps that should be implemented by the data-driven company by 2025. The timeframe looks pretty ambitious, but the goal is not unfeasible given the recent drive to include artificial intelligence in as many company activities as possible. Data is expected to be omnipresent and effectively used in the decision-making process. In the analysis process, thanks to cloud services and real-time access, financial experts will use virtual and augmented reality to make more comprehensive decisions in a multidimensional space. Thereafter, it is necessary to integrate various databases into a single “360-degree view” to properly understand the needs of banking customers and support them by offering rankings to companies they invest in. The provision of and access to securely kept data should be rapid, and all relevant applications submitted by customers must be quickly processed. Finally, the chief data officers will be able to monetize services related to data manipulation (McKinsey Digital, 2022).

The generative AI will improve automation, especially in tedious processes in banks, and it will also augment productivity and accelerate innovative drives. However, with the broader use of generative AI, humans are still a valuable control factor, because these systems are capable of producing impeccable reports with erroneous or non-existent sources. The hallucinations based on model predictions rather than reality could be devastating if distributed inadvertently (Agrawal et al., 2022). Then, the outcomes of genuine and nefarious actors’ actions could be indistinguishable.

Banks were the early adopters of robo-advisors, and it is to be expected that natural language processing will allow for more human-like interaction with customers, not only via online chats but also via the phone. AI-assisted robo-advisors reduce confirmation and hindsight biases routinely demonstrated by human financial advisers (Hasan et al., 2024). For the latter option, i.e.

phone conversation, further improvements are still expected, because the latency in responding and the comprehensiveness of provided answers may not be satisfactory. Moreover, if the system were prone to hallucination, then the customer's trust could be destroyed. The use of Robotics Process Automation will reduce the time necessary for completing repetitive and relatively simple tasks in the back office, which will either reduce the workforce or free up employees' time for more complex tasks (Sushant Das et al., 2024).

Banks pay special attention to streamlining credit approvals. At the same time, it is one of the high-risk AI areas to which the EU Act on Artificial Intelligence would pay special attention. The data collection and analytics, as the cornerstone of the AI breakthrough in this domain, or behavioural biometrics cannot be used to exploit the weaknesses or violate the fundamental rights of the applicant, despite the necessity to examine a multitude of applicants' features to secure the repayment of the loan. Another problem is the accuracy of machine learning, which is a significant concern for banking regulators. One of the solutions could be to refocus from accuracy to optimizing some other specific functions that will boost bank profits (Gramespacher and Posth, 2021).

Ultimately, the banks have to increase investments in preventing fraud and improving the security of collected datasets. The use of AI in data-driven banks will tremendously increase the necessity of evaluating, sharing, and storing datasets, thereby creating many points where potential vulnerabilities may be exploited.

The first step is to allow future customers to verify who they are. In bank branches, the photo from the document would be compared with a person standing in front of the clerk, but it has been shown that humans are not good at matching old photos with 3D appearances. Fintech companies, such as Wise or Revolut, request the submission of real-time photos, passports or other documents via the app during online registration, which is a rich source of information for providers. Phone numbers, email accounts, phone switching or

spending patterns are all useful indicators for evaluating clients (Noonan, 2023). Nevertheless, financial institutions have to move a step further and explore the use of behavioral biometrics. In the first half of 2023, frauds committed by CEOs, banks or law enforcement officials in the UK, thanks to voice generation or video deepfakes, amounted to almost 9 per cent out of £580 million (Quinio, 2024). Payment service providers in the UK have a vested interest in improving the safety of apps, because starting from October 2024, the cost of a single app fraud must be equally shared by sending and receiving payment service providers up to the maximum amount of £415 (PSR, 2023). Faked videos can be used for identification purposes via apps, and if these are difficult to recognize by bank verifiers, one has to explore the advancements in technologies that will recognize single persons among many more accurately.

The system may assume the account has been hacked if users demonstrate unusual browsing patterns. It is also possible to embed watermarks in faked videos for verifiers to recognize them, but the problem is that these watermarks could be hacked, and if it is attached to an original video, the entire process is compromised (Quinio, 2024).

So far, banks have been constrained to use physiological biometrics such as fingerprint or face recognition. Still, the next step involves using behavioral biometrics, such as swiping, keystroke dynamics, gait activities, browsing or touching (Kokal et al., 2024). The modalities of the client's behavior represent a rich source, but it also puts this mode of authentication on a collision course with current and future legal acts. If it becomes quite easy for hackers to compromise accounts, financial services must introduce a so-called continuous authentication to prevent interruptions in service provision and reduce harm. The more modalities the authentication process examines, the more effective it is. However, the moment the data are stored in the server, there is a high likelihood for someone to access the private data unlawfully. One of the solutions could be hashing or returning the authentication data to the client without even letting the server know what indices are used during the authentication (Baig et al.,

2023). Personal data can also be protected by de-identification that modifies and obscures information from the public prying eye (Shopon et al., 2021).

5. CONCLUDING REMARKS

Financial institutions demonstrate a comparatively higher level of confidence vis-à-vis other industries in brandishing their ability to deal with cybersecurity threats. The lack of a complex structure allows their IT experts to pinpoint inconsistencies earlier. The move to the data-driven business environment, strongly dominated by AI-assisted platforms, will further reduce this timeframe, optimize timeframes and allow employees to focus on more complex assignments.

Europeans have made a quantum leap by adopting a standalone Artificial Intelligence Act in March 2024, when major economies struggle to attract the best global talent and reinvent themselves as AI innovation hubs. This may be a risky strategy if the industry, entirely in its early stages, perceives the scope of regulation as onerous. As claimed in this paper, non-EU countries have already attempted to poach EU start-up companies, offering them an unbridled research environment and a range of incentives. So far, other countries have decided to expand the discussion or try to regulate the area by amending existing legal acts. Ultimately, supranational standards will be agreed upon, but no one can accurately predict what the AI success story by countries and regions will look like by then. Europe may have an educated talent pool, but China and the US offer better financial support. Finally, many contenders for the third and other positions are on the leaderboard.

The advancement of AI-modified voices, photos and videos allowed malicious actors, even those with minimal technical skills, to launch highly sophisticated attacks on individuals and institutions. Fake personalities or relatives could illicitly access accounts or persuade people to transfer money to them. Wary of these challenges, the UK regulators have already accepted the

legislation that forces sending and receiving payment service providers to cover damages incurred via the application equally.

AI-generated threats are multifaceted, but banks and other financial institutions are urged to take advantage of the latest artificial intelligence advancements. AI will be omnipresent, tasks will be quickly resolved, data will be available for all relevant parties in a quick manner, and productivity will finally increase considerably. However, with the usage of cloud computing and a flexible work environment, financial actors predict that the primary venues for hackers' attacks will come via these two entry points. If authentication is compromised frequently, it is necessary to advance the use of behavioral biometrics while also paying attention to preserving each client's privacy.

REFERENCES

1. Agrawal, A., Ellencweig, B., Sood, R., Tam, M. (2022). Gen AI: A guide for CFOs. McKinsey and Co.
2. Artificial Intelligence Act 2024. P9_TA(2024)0138, European Parliament 2019–2024.
3. Baig, A.F., Eskeland, S., Yang, B. (2023). "Privacy-preserving continuous authentication using behavioral biometrics." *International Journal of Information Security* 22: 1833-1847. <https://doi.org/10.1007/s10207-023-00721-y>
4. Bank of England (2023). Systemic Risk Survey Results- 2023 H2.
5. Brynjolfsson, E, Rock, D., Syverson, C. (2019). "Artificial intelligence and the modern productivity paradox: A clash of expectations and statistics." In Agrawal, A., Gans, J., Goldfarm A. (Eds.), *The economics of artificial intelligence: An agenda*. National Bureau of Economic Research, The University of Chicago Press: Chicago and London.
6. Bridewell (2023). *Cyber security in critical national infrastructure organizations: Financial Services*, Research Report.

7. Cho, R. (2023). "AI's growing carbon footprint." State of the Planet, Columbia Climate School, Climate, Earth and Society.
8. Cornish, C. (2024). "Abu Dhabi in talks to invest in OpenAI chip venture." Financial Times, 15 March, 2024.
9. Crafts, N. (2021). "Artificial intelligence as a general-purpose technology: A historical perspective." Oxford Review of Economic Policy 37 (3): 521–536. <https://doi.org/10.1093/oxrep/grab012>
10. Criddle, C., Cornish, C. (2024). "Rival nations seek to poach top UK and European AI start-ups." Financial Times, 20 March, 2024.
11. De Cremer, D., Garry K. (2021). "AI should augment human intelligence, not replace it." Harvard Business Review, Winter Special Issue: 97–100.
12. Dora, S. (2024). "AI has a large and growing carbon footprint, but there are potential solutions on the horizon." The Conversation.
13. Dou, W.W., Goldstein, I., Ji, Y. (2023). "AI-powered trading, algorithmic collusion, and price efficiency." Jacobs Levy Equity Management Center for Quantitative Financial Research paper. <http://dx.doi.org/10.2139/ssrn.4452704>
14. Gramespacher, T., Posth, J. A. (2021). "Employing Explainable AI to Optimize the Return Target Function of a Loan Portfolio." Frontiers in Artificial Intelligence, 4(June): 1–13. <https://doi.org/10.3389/frai.2021.693022>
15. Haenlein, M., Kaplan, A. (2019). "A brief history of artificial intelligence: On the past, present, and future of artificial intelligence." California Management Review 61 (4): 5–14. <https://doi.org/10.1177/000812561986492>
16. Hasan, Z., Vaz, D., Athota, V.S., Maturin Desire, SS, Pereira, V. (2024). "Can artificial intelligence (AI) manage behavioral biases." Journal of Global Information Management 31 (2): 1–18. doi: 10.4018/JGIM.321728
17. Jaskiewicz, R. (2023). "Unleashing the power of AI in secured finance industry: A journey into change and evolution." The Secured Lender (July-August): 64-65.

18. Kokal, S., Vanamala, M., Dave, R. (2023). “Deep learning and machine learning, better together than apart: A review on biometrics mobile authentication.” *Journal of Cybersecurity and Privacy* 3: 227–258. <https://doi.org/10.3390/jcp30200134>
19. Kumar, S., Choudhury, S. (2022). “Humans, superhumans, and superhumanoids: Debating Stephen Hawking’s doomsday AI forecast.” *AI and Ethics* 3: 975–984. <https://doi.org/10.1007/s43681-022-00213-0>
20. Marani, M., Soltani, M., Bahadori, M., Soleimani, M., Moshayedi, A. (2023). “The role of biometrics in banking: A review.” *AIRO, EAI*. DOI: 10.4108/airo.3676
21. McKinsey Digital (2022). *The data-driven enterprise of 2025*. McKinsey and Company.
22. NIS 2 (2023). Directive (EU) 2022/2555 of the European Parliament and of the Council of 14 December 2022 on measures for a high common level of cybersecurity across the Union, amending Regulation (EU) No 910/2014 and Directive (EU) 2018/1972, and repealing Directive (EU) 2016/1148 (NIS 2 Directive).
23. Noonan, L. (2023). “CEOs relish prospect of AI boost to productivity: Artificial intelligence banks expect the technology to increase efficiency.” *Financial Times*, 18 September 2023.
24. PSR (Payment System Regulator) (2023). Policy statement: Fighting authorized push-payment scams: Final decision, PS 23/4, December.
25. Quinio, A. (2024). “UK banks prepare for deepfake fraud wave.” *Financial Times*, 19 January, 2024.
26. Rabesandratana, T. (2018). “Europe moves to compete in global AI arms race.” *Science* 360 (6388). doi: 10.1126/science.360.6388.474-a.
27. Rawlinson, K. (2015). Microsoft’s Bill Gates insists AI Is a Threat. *BBC News*. Available online: <https://www.bbc.co.uk/news/31047780> (accessed on 15 March 2024).
28. Rubin, H., Manes, A. (2024). “President Biden Signs Executive Order on ‘Safe, Secure and Trustworthy’ Artificial Intelligence.” *The Computer and Internet Lawyer* 41 (2): 13–14. ISSN: 1531–4944

29. Shen, L. (2023). “World news: AI-driven fraud makes news in China.” Wall Street Journal, Eastern ed., NY 5 June, 2023.
30. Sheppard, L.R., Karlen, R., Hunter, A.P., Baliero, L. (2018). “Artificial intelligence and national security: The importance of the AI system.” A Report on the CSIS Defense-Industrial Initiatives Group, Center for Strategic & International Studies.
31. Shmuha, N. A. (2021). “From a ‘race to AI’ to a ‘race to AI regulation’: Regulatory competition for artificial intelligence.” Law, Innovation and Technology 13 (1): 57–84. <https://doi.org/10.1080/17579961.2021.1898300>
32. Shopon, M., Tumpa, S.N., Bhatia, Y., Pavan Kumar, K.N., Gavrilova, M.L. (2021). “Biometric systems de-identification: Current advancement and future directions.” Journal of Cybersecurity and Privacy 1: 470–495. <https://doi.org/10.3390/jcp1030024>
33. Sushant Das, S., Manohar, A., Jawajala, R.P., Dhanalakshmi, K., Jadhav, M., Raje Sisodia, D. (2024). “AI application in personalized marketing and customer engagement in the retail banking industry.” Academy of Marketing Studies Journal 28 (2): 1–10.
34. Venkataramakrishnan, S. (2024). “AI heralds the next generation of financial scams.” The Financial Times, 19 January, 2024.
35. Zvelo (2023). The role of AI in social engineering. Available at: <https://zvelo.com/the-role-of-ai-in-social-engineering> (Accessed on 18 March 2023).

APPLICATION OF ARTIFICIAL INTELLIGENCE (AI) IN BIOTECHNOLOGY AND MEDICINE

Vera Popović¹

Institute of Field and Vegetable Crops, Novi Sad, Republic of Serbia

Jelena Bošković²

Metropolitan University, Belgrade, Republic of Serbia

Aleksandar Filipović³

*University of Montenegro, Faculty of Medicine, Podgorica,
Montenegro*

Marko Popović⁴

“Svetozar Marković” Gymnasium, Novi Sad, Republic of Serbia

Abstract: *Artificial Intelligence (AI) is the creation of intelligent systems that perform tasks requiring human intelligence, such as learning, problem-solving, and decision-making. Humans and AI systems work together. This study summarizes the potential of AI and its application in medicine, agriculture, and biology-based industries. AI in agriculture provides solutions for food security by adapting agricultural management in a changing climate. Extreme temperatures can reduce wheat yields by 6% per °C. Digitalization in agriculture improves the collection and recording of data on soil health. A reservoir of genetic resources for crops and soil is provided in biodiversity ecosystems, which are key for*

1 vera.popovic@ifvcns.ns.ac.rs

2 jelena.boskovic@metropolitan.ac.rs

3 a.filipovic@t-com.me

4 markopop2006@gmail.com

the diversity of micronutrients. Traditional medicine is widely used by 60% of the world's population, and it originates from medicinal plants from wild populations. As the field of AI evolves with more trained algorithms, the potential for its application in epidemiology, studying host-pathogen interactions, and drug design expands. AI relies on digital technology and is applied in several areas of pharmacy, adaptive medicine, gene editing (CRISPR: a new revolution in genetic technology), radiography, image processing, and drug management. AI is used to identify patterns of new drugs, optimize existing therapies, and use an individual's genomic data and other types of health data to develop personalized treatment plans tailored to their specific needs. It is also used for data analysis, e.g., electronic health records and wearable devices, to identify patterns and correlations that may indicate the presence of a particular disease, helping to improve diagnosis accuracy and enable earlier intervention to prevent disease progression, as well as for medical imaging to identify abnormalities and diagnose diseases.

Keywords: *Artificial Intelligence, medicine, biotechnology, digitalization, data analysis.*

INTRODUCTION

Artificial Intelligence (AI) offers transformational potential in all sectors and industries, ranging from supply chain management (Chi et al., 2020; Nissen and Sengupta, 2006) to medicine (Ali et al., 2018; Cepolina and Muscolo, 2014; Mettler et al., 2017; Vang et al., 2015) and automotive (Lugano, 2017). AI provides opportunities to reinvent business models (Duan et al., 2019), change the future of work (Schvartz et al., 2019), improve performance (Wilson, Daugherty, 2018), and enhance human capabilities (Dwivedi et al., 2021; Collins et al., 2021). AI-based tools and applications help automate complex production processes, meeting the rapidly growing demand for pharmaceuticals, industrial-use chemicals, food, and other raw materials based on research of complex

biological systems. Machine learning (ML), a subset of AI, helps predict outcomes by executing vast permutations and combinations of data sets available for drug molecules to determine the best combination without relying on traditional manual methods in the laboratory. Although traditional model-based methods are still useful for analyzing biological data, they lack the capability to leverage the vast amounts of available data, or even big data, to discover insights, predict data behavior, and understand complex relationships between obtained data. Extensive use of big data is becoming increasingly important in biotechnology and bioinformatics as it continues to grow and become available for scientists to analyze worldwide. There are multiple reasons for the rise in interest in artificial intelligence in recent years (Von Krogh, 2018). In the past few decades, there have been tremendous advancements in some of the foundational methods of artificial intelligence (conventional neural networks), many of which have become open-source and therefore available to everyone. AI requires extensive and sophisticated computing, so the decreasing costs of computer hardware and dedicated AI chip designs are more appealing to organizations. The emergence of the COVID-19 coronavirus increased interest in artificial intelligence as people become accustomed to the reduced human element at all levels of society and the increased use of automation (Coombs, 2020; Sipior, 2020). AI and deep neural network designs can efficiently analyze genomic data to determine the genetic basis of traits and uncover genetic markers associated with specific traits. AI can help decipher complex relationships between different pieces of information hidden in data to extract meaningful results for interpretation and practical implementation. AI attracts significant attention due to its capabilities for faster processing of massive data and extracting meaningful information. AI-based digital image processing, drug design, and virtual drug trials could transform medical sciences in the near future (Ghaffar et al., 2023). This study highlights how artificial intelligence and its components can be used in the medical, agricultural, and bio-based industrial sectors to make human life more sustainable.

ARTIFICIAL INTELLIGENCE (AI) IN MEDICAL SCIENCES

A wide range of medical diagnoses is based on the analysis of disease images obtained using high-tech digital devices. The application of artificial intelligence (AI) in evaluating medical images has led to accurate assessments being performed automatically, which in turn has reduced the burden on doctors, minimized errors and diagnosis time, and improved performance in predicting and detecting various diseases. AI techniques based on medical image processing are an essential area of research that uses advanced computer algorithms for prediction, diagnosis, and treatment planning, resulting in a significant impact on decision-making procedures. Machine learning (ML) and deep learning (DL) as advanced AI techniques are two main subfields applied in the healthcare system for diagnosing diseases, drug discovery, and identifying risk factors for patients (Ghaffar et al., 2023). Advances in medical science and biotechnology have opened new avenues for the development of drugs and antibiotics. AI has enormous potential for broad application in the pharmaceutical industry. Using AI, new therapeutic molecules can be discovered based on known target structures. The branch of artificial intelligence known as ML is commonly used in disease diagnostics as it uses diagnostic test results to improve the accuracy of outcomes. AI enables researchers to manage the prediction and analysis of diagnoses, patient responses to treatment, and patient survival. This includes quantitative and predictive epidemiology, precision medicine, and host-pathogen interactions. AI can aid in the detection and diagnosis of diseases and make computer code more understandable for non-technical experts. Predictive epidemiology, individually-based precision medicine, and analysis of host-pathogen interactions are examples of research areas that could benefit from breakthroughs in machine learning and deep learning. These approaches help in diagnosing diseases and identifying individual cases, making more accurate predictions with fewer errors, faster decision-making, and better risk analysis. The increasing number of tissue biomarkers and the complexity of their evaluations significantly promote the use of AI-based techniques. These AI-based biomarkers

assist doctors in diagnostics. More realistic models of complex socio-biological systems are achievable due to knowledge representation and reasoning modeling. ML-based methods can also be used to improve the efficiency and reliability of epidemiological models. Advances in ML have helped develop ten cellular parameter algorithmic models based on programs that can accurately distinguish benign from malignant tumors (Ghaffar et al., 2023).

It is important to consider individual differences in genetics, ecology, and lifestyle in precision medicine. Doctors recognize that an individual's metabolism, physical and physiological characteristics, and genome structure affect how their body responds to drugs. Despite this, a blanket approach that treats all patients with the same drug, regardless of their different conditions, is still used. Thanks to advances in artificial intelligence, a new era of personalized medicine is developing, where pharmaceutical products are tailored to the needs and adaptability of the organism. Although the transition may seem straightforward, it involves a significant amount of data collection, processing, maintenance, and clarification of the obtained data. Millions of analyses and predictions will be involved in the process to identify the best therapeutic approaches for molecules for a specific case. Using this strategy, doctors and clinicians can better predict which treatment and prevention strategies will be most effective for specific patient groups. Researchers could use AI in studies of DNA, RNA, and proteins to better visualize the effects of drug doses on living tissue over time and reorganize signaling networks during therapy. Based on artificial intelligence, IBM Watson helps create an appropriate treatment plan for a patient depending on their medical history and personal data, including genomic structure. An AI-based personalized medicine system will not only reduce treatment costs but also minimize drug side effects in patients. In addition to saving time and improving patient care, AI can also simplify gene editing, radiography, and drug management planning procedures. Electronic health records (EHRs) can be enhanced with evidence-based clinical decision support systems. Machine learning (ML) is based on learning methods and can be divided into three categories: supervised (classification, regression,

and composition), unsupervised (association, clustering, and dimensionality), and reinforcement learning (Wu et al., 2021), Figure 1. AI includes enormous processing capacity (supercomputers), algorithms that can learn at phenomenal speeds (deep learning), and a new strategy that leverages the cognitive talents of the doctors themselves.

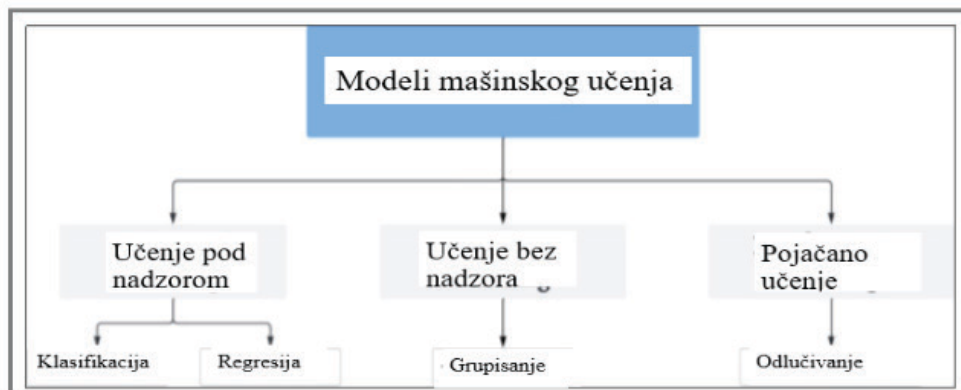


Figure 1. Machine Learning models and main algorithms

This technique contributes to the development of innovative theoretical models of disease pathophysiology and can help predict the major adverse effects of prolonged drug use. It has been found that an AI-based approach is very useful for the early identification, diagnosis, prognosis, and treatment of myopia. In cardiology, dermatology, and oncology, deep learning algorithms show good results in disease diagnostics. Computer algorithms can detect breast cancer metastasis in biopsies of positioned lymph nodes on whole-slide images with an accuracy rate of over 91%, which increased to 99.5% when supplemented with physician data. One proven application of artificial intelligence in risk analysis is the diagnosis of heart disease through cardiovascular imaging. It includes automated monitoring of all deviations from normal conditions based on image processing, myocardial function, and detection and analysis of coronary atherosclerotic plaques (Ghaffar et al., 2023). There are several variables that can be effectively analyzed using artificial intelligence, such as determining conditions that are resistant to certain antibiotics but not others. Such analysis

can support doctors and significantly reduce unnecessary testing and healthcare costs.

One of the most important subfields of AI is machine learning (ML), with essential subfields being neural networks (NN) and deep learning (DL) (Figure 2).



Figure 2. Relationship between AI, ML, NN, and DL approaches

It is important to emphasize the importance of combining these algorithms with medical expertise. New pharmaceutical compounds can be discovered through data analysis using artificial intelligence, reducing the need for clinical trials, and allowing drugs to reach the market faster without compromising their safety. The onset of genetically predisposed diseases may be predicted significantly earlier with the help of AI. Patients will also be able to prevent and treat certain hereditary diseases. One application of artificial intelligence in the pharmaceutical industry is the use of “open target,” a relatively new scientific strategic effort to explore the relationship between drug targets and diseases, as well as how certain genes are related to diseases. SPIDER is another AI technique designed to determine the role of natural products in drug discovery. Quantitative structure-activity relationship (QSAR) studies are particularly useful in creating new effective drugs in a very short period using computer simulation tools. The QSAR model, based on the radial basis function (RBF) artificial neural network (ANN) model, trained using the particle swarm optimization (PSO)

technique, was used in a study to predict pKa values for 74 different types of drugs. Natural language processing (NLP), ML, and robotic process automation are clearly three key areas of advancement for AI in the field of medicine. NLP has recently been used to improve colonoscopy analysis, enhancing the accurate detection of adenomas and polyps. Additionally, the ML approach can be used to predict diseases such as atrial fibrillation and urinary tract infections in certain patient groups using models like the support vector machine (SVM) based on the clinical characteristics of the diseases. Similar initiatives have been used to improve the prognosis of heart disease using heart murmur detection technology. The Food and Drug Administration (FDA) has already approved up to 29 AI-based medical devices and algorithms in various medical science fields (Ghaffar et al., 2023).

The first AI-based model approved by the FDA in the healthcare sector was a diagnostic model based on an autonomous AI system. The model was trained with a diverse set of data samples consisting of individuals of different ages, races, and genders, minimizing the chances of errors across different groups. Several randomized clinical trials (RCTs) have also been conducted to test the efficacy and safety of AI and ML models in clinical practice. In an RCT, the impact of an automated polyp identification algorithm based on deep learning on polyp detection accuracy and adenoma detection rate (ADR) was assessed. In this RCT, consecutive patients were randomly assigned to undergo colonoscopy with or without the aid of an automated polyp identification model, which provided simultaneous optical results and audible alerts upon polyp detection. The results obtained from patients who underwent the AI-based automated detection system outperformed the control groups in ADR and the average number of adenomas and polyps detected per colonoscopy. This automated technology could thus be relevant in treatment regimes and routine practices for improved colon polyp identification due to its high sensitivity, accuracy, and stable results. The introduction of AI systems in medical decision-making has also resulted in cost-effectiveness in overall medical treatment. The pharmaceutical industry will

more comprehensively clarify very complex genetic information with improved AI and ML skills. Clearly, when integrated with ML and NLP, robotic process automation has significant applications and the potential to reshape medical science in the near future. Despite the significant progress made, much more research is needed before AI-based therapy becomes a reality (Lam et al., 2022).

Traditional medicine is widely used by 60% of the world's population, and it originates from medicinal plants from wild populations and cultivation. AI is used in medicine in drug discovery and development: AI is used to identify patterns that help in identifying new drugs and drug targets, as well as optimizing existing therapies. AI is also used in personalized medicine. It uses an individual's genomic data and other types of health data to develop personalized treatment plans tailored to their specific needs, using machine learning algorithms to predict an individual's response to a particular treatment and identify potential adverse reactions. AI is also used in disease diagnosis and prediction. It is used to analyze data, such as electronic health records and wearable devices, to identify patterns and correlations that may indicate the presence of a particular disease, helping to improve diagnosis accuracy and enable earlier intervention to prevent disease progression. AI is used in the analysis of biomedical images. It is used for medical images, such as CT scans and MRI images, to identify abnormalities and diagnose diseases, using deep learning algorithms for automatic segmentation and classification of structures in medical images. The adoption of AI in pharmacological research raises ethical considerations. Ensuring data privacy and security, addressing bias and algorithm transparency, obtaining informed consent, and maintaining human oversight in decision-making are key ethical issues. Responsible AI application requires robust frameworks and regulations. The future of AI in pharmacological research is promising, with integration with new technologies such as genomics, proteomics, and metabolomics offering the potential for personalized medicine and targeted therapies. Collaboration between academia, industry, and regulatory bodies is essential for the ethical application of artificial intelligence in drug discovery and development. Continuous research and development of AI techniques and comprehensive training programs will

empower scientists and healthcare professionals to fully exploit the potential of artificial intelligence, leading to improved patient outcomes and innovative pharmacological interventions (Singh et al., 2023).

ARTIFICIAL INTELLIGENCE (AI) IN AGRICULTURAL BIOTECHNOLOGY

Facial recognition, predicting cancer in tissues, and metabolic flux analysis are just a few examples of significant advancements achieved with AI approaches, and there is potential for a similar revolution in agriculture. According to a report published by the Food and Agriculture Organization (FAO) of the United Nations, the global population will reach over 9 billion by 2050. This demographic growth will ultimately pressure the agricultural sector's ability to provide sufficient food for the human population. To feed the growing global population and enhance the national economy, agriculture is a strategic sector. It is a significant source of income for many countries worldwide (Popović et al., 2010; 2011; 2012; 2013; 2014; 2016; 2018; 2020; 2021; 2022; Bošković et al., 2023a; 2023b; 2023c; 2023d). Agriculture occupies about 38% of the total land area of the planet. Most agricultural activities are now manual, and agriculture can greatly benefit from automation in terms of yield and resource input. Implementing technological discoveries in agriculture can contribute to changing rural economies and the livelihoods of agricultural producers. Agricultural techniques are generally designed to overcome various obstacles, including pest pandemics, inefficient pesticide and fertilizer use, weeds, drought, and lack of adequate irrigation systems, inefficient harvesting, storage, and ultimately marketing. The agricultural sector could be transformed by AI intervention in areas such as land management, water needs assessment, precise mapping of fertilizer, pesticide, insecticide, and herbicide needs, yield prediction, and overall crop management. With advancements in AI-based technology, drones and robots are used to improve real-time crop monitoring, harvesting, and subsequent processing. Biotech companies are currently using AI and ML techniques to design and train autonomous robots capable of performing

key agricultural activities such as crop harvesting much faster than traditional methods. Data collected by drones are processed and evaluated using deep learning and computer vision techniques. Machine learning approaches help access and predict a wide range of environmental variables that affect agricultural production, such as weather fluctuations and meteorological disasters in certain world regions. As mentioned, AI-based solutions in the agricultural industry help improve efficiency and control numerous aspects, such as crop yield, soil profile, crop irrigation, content detection, weeding, and crop monitoring (Popović et al., 2019; Stevanović et al., 2019; 2023; 2024).

Traditional and older examinations of morphological characteristics are time-consuming, prone to errors, and very expensive. Machine vision methods can be easily applied in agricultural practices to speed up and simplify procedures while improving precision and accuracy. Identifying and selecting improved hybrids and varieties can accelerate and facilitate the process using automated, non-invasive, rapid evaluation of various plant characteristics through high-throughput phenotyping methods. Thanks to AI tools, intelligent bundles and drone technology can now be used for several agricultural activities (Ljubičić et al., 2023). Recent developments in AI-based DU and ML algorithm design for agricultural product price estimation can enable farmers to achieve higher profits for their work and investment.

The Internet of Things (IoT) is used to provide easier living, safety, increased productivity, monitoring, and resource optimization in various industries. Agriculture is one of them, where IoT and robots are used before and after the cultivation process, from preparing the land for cultivation to supplying the consumer market. These domains include crop monitoring, smart irrigation, pest monitoring, smart pest control, harvesting, and safe supply to the consumer market while maintaining the quality and integrity of the final product. Thus, new automated methods have been introduced. These new methods have met food needs and enabled the employment of billions of people. Artificial intelligence has brought a revolution in agriculture. This technology has protected crop yields

from various factors such as climate change, population growth, employment issues, and food security problems (Kumar et al., 2022).

For efficient irrigation, artificial neural networks, fuzzy logic, and metaheuristic algorithms have recently been developed. According to a recent study, a convolutional neural network (CNN), which considers several environmental variables, is one of the most reliable ML algorithms for estimating soybean and maize yields. Recent advances in AI-based biosensors for early disease detection in plants, even in asymptomatic plants, have the potential to greatly minimize product loss caused by biotic stresses. AI-based drone technologies, such as EfficientNetV2, designed to detect and classify plant diseases with accuracy and precision of 99.99% and 99.63%, respectively, are among the promising automated technologies for monitoring plant health in a time-saving and cost-effective way. To detect bacterial leaf spots in cultivated crops, an AI hybrid model based on convolutional autoencoder (CAE) and CNN achieved 99.35% and 99.38% in training and testing periods (Bhardwaj et al., 2022).

Due to his initial observations and conceptualization of smart machines, Alan Turing is considered the father of artificial intelligence and modern computer science. He was an early advocate of the theory that the human brain essentially functions as a digital computer (Akman and Blackburn, 2000). He pioneered the experiment known as the “Turing Test,” which became a key moment in AI development. His work, titled “Computing Machinery and Intelligence,” dealt with the possibility of a non-living computer thinking like a human and was a landmark in this field (Turing, 1950). Several other significant events paved the way for the development of the AI we see today (Figure 3, Bhardwaj et al., 2022). The use of AI can facilitate the identification of potential targets in large genomic data sets for genetic manipulation and design efficient synthetic promoters to improve plant agronomic traits. Increasing needs for smart agriculture have led to significant advancements in AI-based agricultural prediction and forecasting, greatly improving crop productivity. A similar attempt was made in a recent study
54 where image data sets were analyzed using AI algorithms, namely ANN and

genetic algorithm (GA)-based platforms, to predict crop yields in an optimized way.



Figure 3. Timeline of Artificial Intelligence development

During the training period, the model achieved a maximum validation accuracy of 98.19%, while a maximum accuracy of 97.75% was achieved during the testing period. This model worked efficiently with limited resources and less data, providing optimal results. In another significant study, a new methodology was proposed for predicting agricultural yield in greenhouse crops using recurrent neural network (RNN) and temporal convolutional network (TCN) algorithms. Based on previous environmental and production data, this approach can be used to more accurately estimate greenhouse crop yields than standard ML and DL methods. This experimental research also highlighted the key importance of previous yield data sets in accurately predicting future crop productivity. Several million individuals in developing countries benefited from the Green Revolution by preventing and combining high-yield crops, synthetic fertilizers, and water. However, due to the widespread misuse of herbicides, pesticides, and fertilizers, the Green Revolution cannot be considered entirely “green.” Certain high-yield crop approaches typically require large amounts of agrochemicals and water. AI-based approaches are being developed to reduce reliance on harmful agrochemicals and achieve sustainability in agriculture. For optimizing agricultural resources, a remote sensing management system (RSCS)

has been developed. This methodology uses AI and ML technology to improve ecological sustainability while encouraging the planning of new agricultural products. When analyzed with other techniques, the findings revealed that RSCS demonstrated the highest accuracy, performance, data transfer speed, productivity, irrigation management, and carbon dioxide release ratio. AI models have the potential to manage agricultural products and productivity in a “green” way. In another study, a smart sprayer based on AI and machine vision was developed to spray herbicides specifically on certain weeds, thereby reducing excessive herbicide use for weed control and environmental pollution. This sophisticated technology combines state-of-the-art weed detection concepts, a unique fast and precise spraying method, and a weed mapping model with 71% and 78% precision and recall. Due to limited data collection techniques and the lack of integration of various data sources, data collection from agricultural regions related to soil hydration, crop quality, or insect infestation often depends on manual analysis (Bhardwaj et al., 2022). As mentioned, AI-based solutions in the agricultural industry help improve efficiency and control numerous aspects such as crop yields, soil profiles, crop irrigation, content detection, weeding, and crop monitoring (Talaviya et al., 2020; Kim, Gilley 2008), Figure 4.

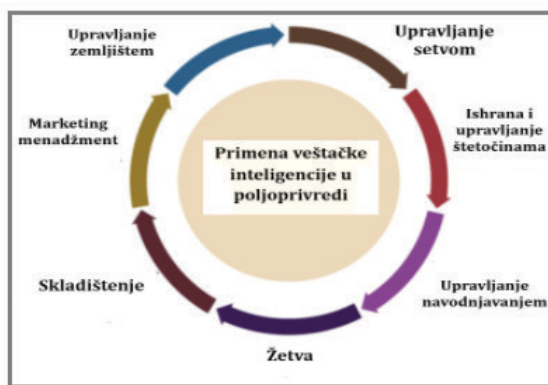


Figure 4. Areas in agriculture where Artificial Intelligence (AI) has a positive impact

56 | Sophisticated AI models reduce the need for agrochemicals. In this regard, the significant NaLamKI action plan, which aims to develop open AI-based

software that could significantly assist the agricultural industry, received financial support from the German government. This plan seeks to develop data sets by combining information from various sensors to optimize different agricultural practices using AI and ML technologies. Similar government initiatives are needed in large numbers to help agricultural producers adapt to artificial intelligence on a larger scale.

In agriculture, integrating precise characteristics based on omics technology data can help identify critical traits involved in stress tolerance and acclimatization mechanisms, as well as contribute to the development of climate-resistant crops. Farmers will be able to generate more production with fewer resources, increase the quality of their production, and ensure faster market time for their harvested crops thanks to the adaptation of AI-based technology. While first-generation AI can be used in omics data research and classification, it is designed to handle specific issues related to single omics data sets without integrating data from other modalities. In agricultural biotechnology, next-generation AI is essentially envisioned to dynamically improve and process large multi-omics data sets in addition to predicting the cultivation values of complex traits under various environmental conditions (Harfouche et al., 2019).

ARTIFICIAL INTELLIGENCE (AI) AND INDUSTRIAL BIOTECHNOLOGY

Industrial biotechnology, known as white biotechnology, is the modern application of biotechnology for the sustainable processing and production of goods, chemicals, and fuels from renewable sources using living cells and their enzymes. The demand for industrial chemicals, pharmaceuticals, food chemicals, and other biochemistry-related raw materials has dramatically increased over the past decade. Machine learning (ML) and AI-based technologies can help design new pharmaceutical products and identify their efficacy and adverse effects before actual production, significantly reducing the time needed to bring drugs from the laboratory to the market for humans. Microorganisms and plant/animal cells are used in biotechnological processing to produce products in various

sectors, including drugs, pharmaceuticals, food, animal feed, disinfectants, cellulose, and textiles. To detect disruptions, optimize machines for efficient production, and improve product quality, the Internet of Things (IoT), ML, and AI can be effectively used. AI-based computer models are becoming increasingly widespread, and robotics and machine learning can be used to develop the best optimal conditions for strain growth and the extent to which valuable products can be obtained. For example, AI-based approaches or response surface methodology (RSM) have been used in the high production of amylase from *Rhizopus microsporus*, using various agro-industrial wastes for optimal experimental designs. Similarly, AI algorithms such as artificial neural networks (ANN) and genetic algorithms (GA) have been integrated to optimize fermentation media for *glucansucrase* production from *Leuconostoc dextranicum*. The integrated ANN-GA model predicted a 6% increase in glucansucrase activity compared to a regression-based prediction approach. The application of the integrated ANN-GA model for optimizing cellulase production from *Trichoderma stromaticum* in solid-state fermentation has been noted, achieving a 31.58-fold increase in cellulase production after optimization with the AI model (Singh et al., 2008).

AI-based technologies have also been used to scale up and optimize bioprocesses for enzyme production on a pilot scale. In one study, a cost-effective method was conducted to scale up the synthesis of extracellular laccase from *Staphylococcus arlettae* using tea waste. RSM and ANN, along with GA, were two successive statistical methods used to increase enzyme production, resulting in a sixteen-fold increase in enzyme yield. Also, some phase-expert system-based visual intelligence models are capable of monitoring pilot-scale wastewater treatment plants. Biofuel is one of the most important bioproducts whose industrial production process can be improved using ML and AI for maximum efficiency. In the bioenergy sector, AI-based approaches are used to predict biomass properties, bioenergy end-use, and the bioenergy supply chain. An integrated ANN-Taguchi method model has been developed to predict and maximize biofuel production through torrefaction and pyrolysis. Optimization and design of experimental factors were carried out using the Taguchi method,

which led to achieving maximum biofuel yield, while ANN showed linear regression prediction for biofuels (Bhardwaj et al., 2022). Integrated ANN-GA models have been used in modeling and optimizing the methanolysis process of waste peanut shells for biofuel production. This demonstrates that the integrated ANN-GA model has better optimization potential than the RSM model alone. ML-based bioprocess models have been constructed using AI-based methods such as ANN, CNN, long short-term memory networks (LSTM), k-nearest neighbors (kNN), and random forests (RF) to predict carbohydrate accumulation in cyanobacteria biomass grown in wastewater for biofuel production. The best results for system dynamics approximation were achieved with 1D-CNN with a mean squared error of 0.0028. Textiles, new chemicals, and biodegradable biopolymer synthesis could benefit from similar processes. It can be used to assist in developing synthesis techniques for such biochemicals that produce the highest yield with the least amount of resources. Additionally, AI could help predict market demand for drugs or chemicals in real-time. AI and ML have also aided metabolite production. Systematic metabolic engineering is a process that helps quickly produce high-performance microbial strains for long-term chemical and mineral production. The increasing availability of large biological data, such as omics data, has resulted in the application of ML techniques at various stages of metabolic engineering systems, such as host strain selection, metabolic pathway reconstruction, metabolic flux optimization, and fermentation. Various machine learning algorithms, including deep learning, have facilitated the optimization of bioprocess parameters and the exploration of a larger metabolic space associated with the biosynthesis of the target bioproduct. This trend also influences biotech companies to adopt ML techniques more frequently in designing their production systems and platform technologies. In the brewing industry, artificial intelligence has shown promising potential to overcome fundamental shortcomings and enhance production through knowledge accumulation and automated control. In one study, AI models were constructed using aroma profiles and spectroscopic data obtained from commercial alcohol to assess beer quality and aroma traits. Intelligent models resulted in highly accurate predictions for six main

beer aromas. ANN model-based technologies have also been developed to assess the presence of various chemicals in beer, such as ethanol, methane, carbon monoxide, hydrogen sulfide, ammonia, and so on. The study involved developing a computer program that simulated the operation of a highly adaptable multilayer perceptron network, which could predict changes in white wine fermentation parameters using data from previous experiments. This study provided a convenient approach to digitizing the brewing process, allowing it to acclimate to other intelligent, knowledge-based frameworks. Another study led to the development of an innovative knowledge-based approach for controlling alcohol fermentation used in white wine production. The primary sources of information used in developing the AI model were various case studies and experimental results, as well as knowledge obtained from brewery experts regarding different parameters related to optimizing and controlling the overall process. Using software for monitoring, regulating, and collecting data from the fermentation bioreactor, an application for automated process control was developed (Gonzalez Viejo et al., 2019, Figure 5). More details about the robotic pourer and computer vision analysis can be found in the work of Gonzalez Viejo et al. (2016).



Figure 5. Equipment for assessing physical measurements of beer; video recorded for beer analysis using computer vision algorithms

Further incorporation of control systems, processes, and innovative advancements can be greatly facilitated by such AI models, thereby supporting sustainable development.

CHALLENGES AND LIMITATIONS OF USING ARTIFICIAL INTELLIGENCE (AI)

Despite its enormous potential, AI-based technologies have yet to enter everyday practice. AI models can enhance the accessibility of various biological sectors but can also exacerbate existing disparities. Since AI models rely heavily on the datasets on which they are developed, as well as the labels associated with them, biases towards underrepresented learning algorithms can be amplified. Several factors must be considered to properly assess the resilience of some deep neural networks. Metadata must be created, retrieved, and cleaned for model development. Programs should be further designed and evaluated under the supervision of field experts to analyze and correct inaccuracies made in practice. Despite significant advances in the design of AI and ML-based models in recent years, few have been integrated into healthcare, and many opportunities for adopting these models for everyday use remain untapped. CNNs, for example, were initially used in study designs starting in 2015, primarily on dental X-rays, and the first clinical uses of these tools only recently emerged. The unavailability of clinical data due to organizational policies, insufficient reproducibility in data set processing and outcome assessment, and ongoing concerns about accountability and transparency towards patients remain the most common barriers to adopting AI in routine medical and dental practice. Several models have been reported to be inaccurate in predicting clinical diagnoses. An AI algorithm has been developed to diagnose and classify chest X-rays using NLP for radiology records. These classifications were later used in training a deep learning network to detect abnormalities in images, with a particular focus on identifying pneumothorax. After a detailed review, the presence of chest tubes in most reports identified as pneumothorax raised the question that the algorithm was recognizing chest tubes instead of pneumothorax as intended

(Liong-Rung et al., 2022, 893208). Due to different marginalizations in training datasets, modeling sensitivity to genetic diseases is also predisposed to varying performances among different demographic groups.

Additionally, the amount of computing time required for ML approaches may vary depending on the data size. Using AI-based approaches will not always lead to improvised categorization or better prediction than current methods. AI is a tool that should be used in the appropriate context to address a relevant issue or solve a significant problem. Similarly, in other biological areas such as agriculture, automating practices using AI and machine learning approaches has great potential for sustainable agriculture. However, in the agricultural sector, data collection, analysis, and usage for productivity present numerous challenges. Data privacy and security are two major challenges that farmers must address to survive in the digital age. In most cases, they are not informed about data collection, usage, and, more importantly, the purposes for which their personal data are used. Data mining enables corporations to rely on individuals to collect massive amounts of agricultural data, which may be sufficient to develop and evaluate behavioral and psychiatric profiles of subjects. To prevent data misuse, agricultural producers require assurance that their information will be used to generate innovative ideas and agricultural solutions, not to gain a competitive advantage. As mentioned, AI-based drone technology has emerged as a very efficient approach in agriculture. However, drones, especially those equipped with high-resolution lenses, infrared cameras, competent programs, and sensors, are very expensive for small agricultural producers. To operate drones, authorization is required following their operational and regulatory provisions of land law. Also, weather has a huge impact on drone operations. Traditional data mining methodologies were primarily developed for relational datasets; however, they are not entirely adequate for geographically dispersed data. To develop agriculture with AI-based technologies, innovative data mining approaches are needed. In the industrial biotechnology sector, establishing defined and sustainable protocols for algorithm adoption and data set size assessment remains a significant challenge. Designing such protocols would require a

thorough knowledge of the effects/efficacy of different algorithms and training datasets to address numerous bioindustry challenges. Increased availability, good documentation, and superior data collection methods are still needed for the development, operation, and optimization of bioenergy systems and bioreactor design. In some AI models, when the input is inadequate, especially for large datasets, the algorithm may instruct each variable as a separate instance instead of learning the information, resulting in errors and lower training efficiency. Additionally, numerous systems represented by ANN are often penalized due to black-box characteristics. Nevertheless, the lack of comparative work in different AI-ML designs makes it challenging to present a clear direction for future studies or practical applications. Challenges still need to be overcome, including inefficient data integration arising from diverse datasets, including candidate data, metadata, processed data, raw data, and the lack of an appropriate set of skills and expertise related to the subject. In this context, it is essential to overcome these uncertainties using new AI algorithms to achieve fundamental compliance between expected outcomes and empirical studies (Holzinger et al., 2023, 24). Therefore, more comprehensive datasets and relative studies are needed to develop AI and ML-based models for real-time monitoring and control of bioreactors and bioprocesses.

CONCLUSION

One of the great achievements we have witnessed in the era of Industry 4.0 is the ability of machines to replicate the capacities of living systems, especially human intelligence. The potential of artificial intelligence could be harnessed in the biological world, including medical research, agriculture, and biology-based industries, for sustainable living. Early prediction and identification of diseases and their precise treatment based on personalized medicine, even when diseases are in asymptomatic states, are examples of key areas of medical science that could benefit from AI. This would save millions of lives and reduce medical costs. Besides the medical field, efficient AI-based algorithms and programs have recently been developed to provide efficient resources and results in

agriculture, a practice known as precision agriculture. Agricultural practices such as land management, water needs analysis, fertilizer needs modeling, pesticides, insecticides, herbicides, yield projection, and overall crop management can be revolutionized by AI intervention. This would meet the growing global demand for food. AI-based programs and computer models have proven to be very effective in optimizing suitable conditions for obtaining the maximum desired product, whether for agricultural, medical, biotechnological, or everyday use, with minimal costs. Efficient production of bioenzymes is just one such success, and it is easy to imagine how the biotechnology industry will be transformed by the application of artificial intelligence, which will help reduce production costs, one of the biggest challenges facing the industry today. Traditional medicine is used by 60% of the world's population and originates from medicinal plants from wild populations. AI is used in medicine in drug discovery and development: to identify patterns that help in identifying new drugs and drug targets, as well as to optimize existing therapies, then in personalized medicine, in disease diagnosis and prediction, in biomedical image analysis, to identify abnormalities and diagnose diseases, using deep learning algorithms for automatic segmentation and classification of structures in medical images.

More than 90% of technological innovations in human history occurred in the 20th century, and it is expected that 99% of innovations will happen in the 21st century. The reason for this is seen in the explosion of artificial intelligence in the middle of the century and the expected emergence of artificial superintelligence in the second half of the 21st century.

We cannot exclude the possibility that the artificial intelligence we have created, driven solely by its own interests, will one day turn against humans if we do not regulate its development in time and incorporate empathy and respect for humans and all living beings on the planet as mandatory elements of the system.

Acknowledgment: The research presented in this paper is part of a project funded by the Ministry of Education, Science and Technological Development of

the Republic of Serbia, project number: 451-03-66/2024-03/200032. All authors contributed equally to this work.

REFERENCES

1. Akman, V., Blackburn, P. 2000. Editorial: Alan Turing and Artificial Intelligence. *J. Logic Lang. Inf.* 9:391–395. <https://doi.org/10.1023/A:1008389623883>.
2. Bhardwaj, A, Kishore, S, Pandey, D.K. 2022. Artificial Intelligence in Biological Sciences. *Life -Basel.* 12(9):1430. <https://doi.org/10.3390/life12091430>
3. Bošković, J., Popović, V., Mladenović, J., Stevanović, A., Ristić, V., Maksin, M., Jovanov, D. 2023a. The future of smart agricultural production through applied information technologies. *IRASA International Scientific Conference Science, Education, Technology and Innovation, SETI V, 14.10.2023, Belgrade*, p. 3-35.
4. Bošković, J., Mladenović, J., Popović, V., Stevanović, A., Ristić, V. 2023b. Significance of new plant breeding technologies for sustainable agriculture and food security. *IRASA International Scientific Conf. Science, Education, Technology and Innovation, SETI V, 14.10.2023. Belgrade.* p.150-180. ISBN 978-86-81512-11-1
5. Bošković, J., Mladenović, J., Ristić, V., Burić, M., Popović, V. 2023c. Genetic approach and explanation of intelligence. *IRASA Intern. Scientific Conf. Science, Education, Technology and Innovation, SETI V, 14.10.2023. Belgrade.* p.113-140.
6. Bošković, J., Mladenović, J., Popović, V., Ristić, V., Stevanović, A., Šarčević-Teodosijević, Lj. 2023d. Specific reactions of plants to abiotic stresses with physiological biochemical and molecular perspectives. *IRASA International Scientific Conference Science, Education, Technology and Innovation, SETI V 2023, 14.10.2023 Belgrade.* p. 369-389.
7. Chi, M., Huang, R., George, J.F. 2020. Collaboration in demand-driven supply chain: Based on a perspective of governance and

- IT-business strategic alignment. *International Journal of Information Management*, 52: 102062.
8. Collins, C., Dennehy, D., Conboy, K., Mikalef, P. 2021. Artificial intelligence in information systems research: A systematic literature review and research agenda. *International Journal of Information Management*. 60: 102383. <https://doi.org/10.1016/j.ijinfomgt.2021.102383>
 9. Gonzalez-Viejo, C., Torrico, D.D., Dunshea, F.R., Fuentes, S. 2019. Development of Artificial Neural Network Models to Assess Beer Acceptability Based on Sensory Properties Using a Robotic Pourer: A Comparative Model Approach to Achieve an Artificial Intelligence System. *Beverages*. 5(2):33. <https://doi.org/10.3390/beverages5020033>
 10. Gonzalez Viejo, C., Fuentes, S., Li, G., Collmann, R., Condé, B., Torrico, D. 2016. Development of a robotic pourer constructed with ubiquitous materials, open hardware and sensors to assess beer foam quality using computer vision and pattern recognition algorithms: RoboBEER. *Food Res. Int.* 89: 504–513.
 11. Coombs, C. 2020. Will COVID-19 be the tipping point for the Intelligent Automation of work? A review of the debate and implications for research. *International Journal of Information Management*, 55: 102182.
 12. Ghaffar Nia, N., Kaplanoglu, E., Nasab, A. 2023. Evaluation of artificial intelligence techniques in disease diagnosis and prediction. *Discov Artif Intell.* 3(1): 5. <https://doi.org/10.1007/s44163-023-00049-5>.
 13. Duan, Y., Edwards, J.S., Dwivedi, Y.K. 2019. Artificial intelligence for decision making in the era of Big Data – evolution, challenges and research agenda. *International Journal of Information Management*, 48: 63-71.
 14. Dwivedi, Y.K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., Medaglia, R. 2021. Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management*, 57: 101994.

15. Harfouche, A.L., Jacobson, D.A., Kainer, D., Romero, J.C., Harfouche, A.H., Scarascia Mugnozza, G., Moshelion, M., Tuskan, G.A., Keurentjes, J.J.B., Altman, A. 2019. Accelerating Climate Resilient Plant Breeding by Applying Next-Generation Artificial Intelligence. *Trends Biotechnol.* 37:1217–1235. doi: 10.1016/J.TIBTECH.2019.05.007.
16. Holzinger, A., Keiblinger, K., Holub, P., Zatloukal, K., Müller H. 2023. AI for life: Trends in artificial intelligence for biotechnology. *New Biotechnology.* 74: 16-24.
17. Kim, M., Gilley, J.E. 2008. Artificial Neural Network Estimation of Soil Erosion and Nutrient Concentrations in Runoff from Land Application Areas. *Comput. Electron. Agric.* 64:268–275. <https://doi.org/10.1016/J.COMPAG.2008.05.021>
18. Kumar, A., Singh, S.R., Yadav, M.C., Bhuj, B.D., Dhar, S., Pruthi, N.K., Kumar, R., Bajpai, V., Rizwan, M., Jyoti, K., Singh Thapa, R., Kumar, V., Kumar, H., Kumar Mishra, B., Kumar, V., Rajput, A., Singh, A., Kumar, R. 2022. Artificial Intelligence, Internet of Things (Iot) and Smart Agriculture for Sustainable Farming: A Review. *Annals of Plant Sciences.* 11(11): 5512-5564. <http://dx.doi.org/10.21746/aps.2022.11.11.6>
19. Lam, T.Y.T., Cheung, M.F.K., Munro, Y.L., Lim, K.M., Shung, D., Sung, J.J.Y. 2022. Randomized Controlled Trials of Artificial Intelligence in Clinical Practice: Systematic Review. *J Med Internet Res.* 24(8):e37188. <https://doi.org/10.2196/37188>.
20. Liong-Rung, L., Hung-Wen, C., Ming-Yuan H., Shu-Tien, H., Ming-Feng T., Chia-Yu, C, Kuo-Song, C. 2022. Using Artificial Intelligence to Establish Chest X-Ray Image Recognition Model to Assist Crucial Diagnosis in Elder Patients With Dyspnea. *Frontiers in Medicine* 9:893208, DOI: 10.3389/fmed.2022.893208
21. Lugano, G. 2017. Virtual assistants and self-driving cars. *Warsaw: 2017 ITST Proceedings.*
22. Schwartz, R., Dodge, J., Smith, N., Etzioni, O. 2019. Green AI. *Commun. ACM*, 63 (12): 54-63.

23. Ljubičić, N., Popović, V., Kostić, M., Pajić, M., Buđen, M., Gligorević, K., Bižić, M., Crnojević, V. 2023. Multivariate Interaction Analysis of *Zea mays* L. Genotypes Growth Productivity in Different Environmental Conditions. *Plants*. 12 (11): 2165. <https://doi.org/10.3390/plants12112165>
24. Nissen, M.E., Sengupta, K. 2006. Incorporating software agents into supply chains: Experimental investigation with a procurement task. *MIS Quarterly*, 30 (1): 145-166.
25. Popović, V., Glamočlija, Đ., Malešević, M., Ikanović, J., Dražić, G., Spasić, M., Stanković, S. 2010. Genotype specificity in nitrogen nutrition of malting barley. *Genetika*, 43(1):197-204.
26. Popović, V., Glamočlija, Đ., Malešević, M., Vidić, M., Tatić, M., Ikanović, J., Jakšić, S., Spasić, M. 2011. Uticaj folijarne prihrane i tretiranja semena preparatom na bazi Co i Mo na prinose soje. *Zbornik Instituta PKB Agroekonomik*, 25(1-2):117-125.
27. Popović, V., Jakšić, S., Glamočlija, Đ., Đekić, V., Grahovac, N., Mickovski Stefanovic, V. 2012. Variability and correlations between soybean yield and quality components, *Romanian Agricultural Research*, 29:131-138.
28. Popović, V., Miladinović, J., Vidić, M., Mihailović, V., Ikanović, J., Đekić, V., Ilić, A. 2014. Genotype x environment interaction between yield and quality components of soybean [*Glycine max*]. *Agriculture and Forestry*. Podgorica, 60(2): 33-46.
29. Popović, V., Malešević, M., Miladinović, J., Marić, V., Zivanović, Lj. 2013. Effect of Agroecological Factors on Variations in Yield, Protein and Oil Contents in Soybean Grain. *Romanian Agricultural Research*, 30: 241-247.
30. Popović, V., Tatić, M., Sikora, V., Ikanovic, J., Drazic, G., Djukic, V., Mihailovic, B., Filipovic, V., Dozet, G., Jovanovic, Lj., Stevanovic, P. 2016. Variability of Yield and Chemical Composition in Soybean Genotypes Grown Under Different Agro-ecological Conditions of Serbia. *Romanian Agricultural Research*, 33: 29-39.

31. Popović, V., Kolarić, Lj., Živanović, Lj., Ikanović, J., Rajičić, V., Dozet, G., Stevanović, P. 2018. Influence of row spacing on NAR–Net Photosynthesis Productivity of *Glycine max* (L.) Merrill. *Agriculture&Forestry*, 64(1): 159-169. doi.org/10.17707/AgricultForest.64.1.18
32. Popović, V., Vučković, S., Jovović, Z., Ljubičić, N., Kostić, M., Rakašćan, N., Glamočlija-Mladenović, M., Ikanović, J. 2020. Genotype by year interaction effects on soybean morpho-productive traits and biogas production. *Genetika*, Belgrade, 52(3): 1055-1073. <https://doi.org/10.2298/GENSR2003055P>
33. Popović, M.V., Šarčević-Todosijević, Lj., Petrović, B., Ignjatov, M., Popović, B.D., Vukomanović, P., Milošević, D., Filipović, V. 2021. Economic Justification Application of Medicinal Plants in Cosmetic and Pharmacy for the Drugs Discovery. Chapter 3. Ed. Emerald M. Book Title: An Introduction to Medicinal Herbs. *NOVA Science publishers*, USA, p.63-106. p.1-365. <https://doi.org/10.52305/TKAL3430>
34. Popović, V., Burić, M., Mihailović, A., Aćimić-Remiković, M., Vukeljić, N., Batrićević, M., Petrović, B. 2022. Medicinal properties of buckwheat products and honey in compliance with food safety regulatory requirements. *Journal of Agricultural, Food and Environmental Sciences*, 76, 3, 16-24.
35. Singh, A, Majumder, A, Goyal, A. 2008. Artificial intelligence based optimization of exocellular glucansucrase production from *Leuconostoc dextranicum* NRRL B-1146. *Bioresour Technol.* 99(17):8201-6. <https://doi.org/10.1016/j.biortech.2008.03.038>
36. Singh, S, Kumar, R, Payra, S, Singh, SK. 2023. Artificial Intelligence and Machine Learning in Pharmacological Research: Bridging the Gap Between Data and Drug Discovery. *Cureus*. 15(8):e44359. <https://doi.org/10.7759/cureus.44359>
37. Sipior, J.C. 2020. Considerations for development and use of AI in response to COVID-19. *International Journal of Information Management*, 55: 102170.

38. Stevanović, A., Šarčević Todosijević, Lj., Bošković, J., Popović, V., Živanović, Lj. 2019. Organska proizvodnja, genetički modifikovani organizmi i očuvanje biodiverziteta-vodeći izazovi u zaštiti životne sredine. *Naučni skup Održiva primarna poljoprivredna proizvodnja u Srbiji-stanje, mogućnosti, ograničenja i šanse. Bačka Topola*, pp.95-102.
39. Stevanović, A., Bošković, J., Popović, V. 2024. Implications of HAARP system on climate change and sustainable agriculture. *International Multidisciplinary Conference Challenges of Contemporary Higher Education*” - CCHE 20224, Kopaonik January 29th - February 2024. p. 563-570.
40. Stevanović, A., Stevanović, S., Jauković, M., Bošković, J., Popović, V., Ristić, V., Šarčević Todosijević, Lj. 2023. Modern organic agriculture in accordance with global GAP standard and HACCP system. *IRASA International Scientific Conference Science, Education, Technology and Innovation, SETI V 2023, 14.10.2023 Belgrade*, pp. 37.
41. Talaviya, T., Shah, D., Patel, N., Yagnik, H., Shah, M. 2020. Implementation of Artificial Intelligence in Agriculture for Optimisation of Irrigation and Application of Pesticides and Herbicides. *Artif. Intell. Agric.* 4:58–73. <https://doi.org/10.1016/J.AIIA.2020.04.002>.
42. Turing, A.M.I. 1950. Computing Machinery and Intelligence. *Mind*. 59:433–460. <https://doi.org/10.1093/MIND/LIX.236.433>.
43. Wilson, J., Daugherty, P.R. 2018. Collaborative Intelligence: Humans and AI Are Joining Forces. *Harvard Business Review*.
44. Von Krogh, G. 2018. Artificial intelligence in organizations: New opportunities for phenomenon-based theorizing. *Academy of Management Discoveries*, 4 (4): 404-409.
45. Wang, C., Savkin, A., Clout, R., Nguyen, H. 2015. An intelligent robotic hospital bed for safe transportation of critical neurosurgery patients along crowded hospital corridors. *IEEE Transactions on Neural Systems and*

Rehabilitation Engineering, 23 (5): 744-754. <https://doi.org/10.1109/TNSRE.2014.2347377>

46. Wu, X, Liu, X, Zhou, Y. 2022. Proceedings of 2021 chinese intelligent systems conference: review of unsupervised learning techniques in lecture notes in electrical engineering. *Singapore*.

INTEGRATION OF ARTIFICIAL INTELLIGENCE IN PROFESSIONAL ARCHITECTURAL PRACTICE

Semir Poturak, PhD¹

Faculty of Technical Studies in Travnik, University of Travnik

Abstract: *The research and application of Information and Communication Technologies (ICT) in various processes within the Architecture, Engineering, and Construction (AEC) industry have been present for decades. This practice continues with contemporary advancements in the field of Artificial Intelligence (AI), especially through software designed for widespread adoption and consumption by architects and other professionals in the AEC industry. The process of creating “digital twins” of buildings – computable models that simulate design, construction, building performance, and maintenance – requires very strict parametrization of different elements and procedures that constitute the composition of an architectural project as a whole. The concrete parametrization of building elements, in return, requires strict theoretical classifications, categorizations, structuring, and the composition of design methodology as a programmable, computable algorithm. Questions arise, such as: What is the logic behind automated decision-making? How much can we trust an automated computational model? What kind of tendencies or ideologies are embedded in the process of identifying parameters, their connections, and their relation to an architectural project? What is the underlying design methodology or design philosophy upon which an algorithm is created? The purpose of*

this paper is to question the integration of AI and ICT in the actual process of architectural design and to explore realistic possibilities for implementing emerging technologies in the architect's workflow that extend beyond research and experimentation into the domain of architectural practice.

Keywords: *architecture, architectural design, architectural engineering, artificial intelligence, AI, AI tools, CAD, parametric design.*

1. THE CULTURE OF BUILDING DESIGN AND CONSTRUCTION

A building's purpose is to enable the use of spaces in a meaningful, comfortable, affordable, and secure way, and to endure and maintain desired conditions through years of weather changes, natural disasters, and many other unwanted and unexpected scenarios. The idea to design and build a building comes from the need for a durable, stable, functional shelter, storage, temple, or other archetype from the architectural language related to living and working culture. Longevity, durability, and stable existence over long periods of time are essential features that predefine the building design and construction process, making it a relatively expensive investment even in the case of a smaller housing unit or a simple hut. Reliable building materials, structural systems, and general requirements of interior spaces in terms of volume and functionality are some of the conditions for the enduring usability of a building, all of which are relatively costly in terms of resources, knowledge, skills, available machinery, and construction technology. Therefore, reducing costs by reducing time and resources used for design, manual labor, the consumption of building materials, and the resources needed for building operation is an ever-growing tendency in the construction industry, especially with the use of rapidly evolving and developing information and communication technologies. However, despite ongoing efforts and investments in cost reduction, the complexity of building design and construction—marked by the purpose of serving its users and maintaining functionality and security over long periods—makes optimizing

the design, construction, operation, and maintenance process an extremely challenging task.

Since the earliest days of humanity, building various machines, instruments, and tools to further optimize, speed up, and enhance different parts of architectural and construction processes has been a common practice. The idea of achieving a goal in different, more efficient ways to produce higher-quality outputs is a consistent theme throughout the history of problem-solving. Measuring distances, and surfaces of regular and irregular land areas, measuring heights of different points on the terrain, recording precise geometry of lots and parcels for insurance purposes, and the creating of various plans and maps are some of the problems in building design and construction that are solved today with various mechanical and digital inventions.

“Indeed, when culture is introduced at all it is treated as a constraint to creativity. For instance: ‘[d]esigns that are based on known compositions of techniques (practices and styles) from within a culture will appear mundane.’” (Calvo M.C., 1993).

On the contrary, design itself may already include a drive for innovation, improvement, and change. The culture of architectural design, construction, and operation of buildings is embedded within the tendency to change, improve, reform, and transform the way things are thought or done – improvement in design and construction is always present. Reproduction, copying, and replication of existing designs may appear mundane, but every new architectural project has its own context and reality with unique features that affect design decisions and the final appearance of the building. Here again, it is the architect’s responsibility to interpret the conditions and context for architectural design and apply skills and use accessible resources to create a unique architectural solution.

Everlasting optimization and innovation in architectural design and construction within the given environment may be the essential forces that drive and define the general culture, history, and tradition of architectural design

and construction. Optimization of a building's life from the first conceptual designs to a usable constructed building is a continuous process in all cultures of architectural design. The ability to design and construct a building in a short period while preserving overall quality defines the art of architectural design and construction. Being practical and innovative in the use of accessible materials, experimenting with new building materials, using different tools, including various scientific and technological discoveries, reinventing and reusing existing technology – anything that could ease, improve, and speed up the design, construction, and operation process while maintaining building quality can be incorporated into the process of architectural design and construction.

The tendency to automate, ease, and transform the process or parts of the process of building design and construction to make the whole process more efficient is enhanced, especially within the capitalist market where form follows profit. Therefore, there are significant concerns regarding the automation of the design process to preserve the quality of building spaces that adhere to social, psychological, and ethical standards and norms. Automating the process or its parts and reducing the time needed for design does not mean there will be significant changes in the overall quality of the building and the quality of services the building should provide. On the other hand, problems of homelessness and the growing need for housing spaces due to the increase in global population and migrations to urban areas drive architects and other engineers to seek more efficient but also sustainable, long-term solutions.

“Digitization and AI require new machines (robots, automation technology), much higher data speeds (fiber optic cables, routers, higher performance microchips, sensors, data transfer infrastructure), and data storage with larger volumes such as significant higher data capacity (data clouds, IoT, AIoT).” (Weber Lewerenz, 2023). The development of new tools, faster communication, travel and transportation, the apparent ease of use, and mass adoption of technology additionally create new layers of archetypes, new types of buildings, structures, and infrastructure, which again are subjects of

optimization and development (e.g., data centers, control rooms). Investing in research of new ideas to improve existing processes is available only to larger companies and institutions. Research conducted by the Fraunhofer Institute Austria in 2022 shows that a redesign of the educational system is also required since AI and other emerging technologies require “new qualifications, new knowledge, improved transfer of knowledge between research and practice, and a new corporate culture of thinking and openness towards innovation.” (Weber-Lewerenz, 2023). Even though the Architecture, Engineering, and Construction (AEC) industry is considered to be the least digitized industry, the AI valuation in the construction market was US\$429.20 million in 2018 and is expected to grow to US\$4.51 billion in 2026. (Rafasanjani, Nabizadeh, 2023).

2. ARCHITECTURAL DESIGN

“Conceptual design continues into design phases, the construction phase, and throughout the life of the building.” (Pena M. L. C. et al, 2021).

In this paper, the focus is on the architectural design process as the initial phase towards the creation of the building, with building construction and building maintenance and operation as consequent stages of building life. The architect’s responsibility is to define in detail the future building and, in this way, determine how the building will be constructed and operated. In essence, architectural design must incorporate detailed information about how the building will be constructed and how it will be used and maintained. The architectural project contains information about all stages of the future building’s life and must provide instructions on how to build and equip an enduring and functional building that is also aesthetically pleasing.

The process of architectural design, therefore, involves projecting different future events and scenarios regarding construction, usage, and maintenance of the building and incorporating them into the initial design – architectural

plan – architectural project – instructions for the construction of the building according to the specified function. In this way, the architectural building design process is heavily influenced by available and accessible construction, maintenance, and operational building technology. Accessibility and affordability of preferred structural systems and components, various devices and installation systems, insulation, and quality and durable building materials are also some of the conditions that shape the architectural design process.

The architectural design process can be represented as a finite linear vector representing time in the sense that there are consequent phases as the project develops. The architect first collects and organizes information relevant to the project – in the second step, the information model is translated into a geometrical, conceptual model. The conceptual model is then fine-tuned and eventually presented as a set of documents that prove its legality and also serve as an instruction for construction. However, the architectural design is not a linear process but has the form of a multidimensional, multilayered model – a virtual (cognitive) model where the architect from the very beginning deals with the totality of available information (constraints, access, resources) and defined requirements of future space to be designed.

Designing complex systems such as buildings is rarely done on the first try; rather, it is a process of exploring different solutions and incorporating feedback that, in return, becomes the base for new solutions and proposals. Architectural design is a process that goes back and forth between the architect (who proposes solutions according to known information) and different stakeholders (client, users of the space, neighbors) until a consensus is achieved. Based on the given information, the architect explores future possibilities to design the building by exploring different architectural solutions – different geometries – and how they interact with given constraints, potentials, and features of the given location.

Every architectural project starts with existing constraints that already predetermine possible building design outcomes. The exact geographical location, geometrical boundaries of the construction site and the terrain, specific user needs, and functional requirements are just some of the constraints that narrow the number

of possible reasonable design outcomes from the very start of the design process.

The architect's trip to the building site, visiting the site, and understanding and comprehending the environment and context of the future building are some of the very important steps in the architectural design process. Physical presence at the location for which the building is to be designed provides the architect with unique insights into the existing qualities, issues, and potentials that should be incorporated into the project – sometimes the existing conditions might be the crucial force that actually shapes the design.

Designing building extensions or adding additional components to an existing building is yet another common challenge for architects. In these cases, there are many more constraints, and the architect has to adhere to the existing geometry, style, and functions in a strict way to design a functional building extension. This task is extremely challenging when renovation, reconstruction, or similar works are done on historical buildings.

In theory, instead of building the extension, removing a part of the building may solve the problem. There may be instances where not designing and not constructing the building may be a better solution than having a building designed and constructed. It may happen that a certain function does not fit a specific location because of socio-economic, urban, infrastructure, or other factors, and no design tool, optimization, or simulation software could eventually help produce a successful design if the more general urban conditions are not met.

The meaning of culture in the context of contemporary building design and construction is phantasmagoric in any sense beyond being practical, functional, sustainable, efficient, and durable – which is the essence of a building. Aesthetics in the domain of computation become a manifestation of the architect's ability to organize, manage, sort, shape, and combine different information into the geometry of architectural language – aesthetic value emerges from the overall composition of spaces and forms that anticipate, host, welcome, and perform given functions. With the use of different optimization software in the initial stages of design, especially in terms of optimizing natural daylight exposure,

passive heating, and shadowing, beauty no longer belongs solely in the eye of the beholder. In this way, beauty appears as a surplus product of the specific management of various building elements and layers into a coherent, unifying physical structure, in a unique location, in a unique moment of time, combining past experiences, contemporary tendencies, and visions of the future.

3. BUILDING AS AN INFORMATION MODEL

“Development of computer-based assistants must be grounded in a larger philosophical discourse which places it in a framework for addressing social concerns and providing validation for daily life.” (Calvo M.C., 1993).

A philosophical approach to architectural design, in the broadest sense, must incorporate a number of different dimensions, perceptions, and realities. It is a generic feature of every architectural project to include information from various users and participants affected by the act of architectural design, engineering, and construction. This also includes information about the location, legal situation, climate, context and neighborhood, living culture, building culture, and available building and construction technology.

Architectural design is partially a task of managing information coming from different disciplines such as psychology (needs of individuals in terms of mental health), sociology (spaces for human interaction, accommodation of different human relationships), physics (the movement of air through the building, sound transfers, structural capabilities of materials in different shapes), and chemistry (reaction of building materials in different conditions). The architect juggles all of this information (and much more) from different sources and transforms it into a meaningful geometrical shape that represents the building.

To produce meaningful geometry and successful building design, the architect first has to gather and manage information relevant to the design of the future building. The architect acts as an information manager, managing different information and producing geometry as a response to that information.

The information the architect has about the project in general before stepping into any sort of design is crucial for the design outcome. With access to new information, the design changes accordingly. It is the architect's responsibility to search, gather, explore, collect, manage, and sort information and determine its relevancy, as well as the information sources related to the project, to get the overall "picture" of the reality of the architectural challenge as accurately as possible.

The more information the architect has about the existing, present context, situation, location, social context, habits, and culture attached to a specific location, and the opinions of different stakeholders, as well as about the new building and its function, requirements, and intentions, the greater the chance for the building to be designed to last and endure for a long period. The more access the architect has to the "objective reality" of the context in which the architectural project will be developed, the greater the chance that the future building will be able to address real-life problems, challenges, and situations that it will encounter after being constructed.

The architect's task is also to collaborate with other engineers (mechanical, plumbing, HVAC, renewable energy, etc.) and incorporate related systems into the overall building design. With the support of a structural engineer, the architect defines and optimizes the structural system of the building, including the exact dimensions of building elements such as beams, columns, slabs, and foundations. Eventually, all related information is used to define and redefine the geometry of the building, incorporating what is required and fitting within the boundaries and constraints. The architect's responsibility is to invent a building, a geometrical shape with physical properties, that fits within legal, physical, and natural constraints, but also to fit the limited resources, time, accessibility to materials, and construction technology.

4. CONCEPT

In the face of a situation that eventually requires architectural intervention, the architect's actual response or output is a more or less geometrical model

based on the assumed distribution of different functions, circulation, structural features, and costs of construction and maintenance. The overall geometry – dominant volumetric – representative, encapsulating shape of the building – the concept – is one of the infinite possible geometrical shapes as an answer to the given architectural challenge. It is a result of concurrent information about the project. With new information, different concepts may be found to be more appropriate, or new geometries can open different questions and perspectives that could then be used to reconfigure existing or create new concepts.

The conceptual phase of architectural design – referred to as the concept – this terminology is mostly used to describe the process of essentially brainstorming by an architect or a team of architects. The conceptual phase refers to the “play” of the architect to envision, imagine, and project a future state of a specific location and its context as a whole by “playing” with different information and geometries.

The conceptual phase is referred to as “The initial form,” “Building’s shape,” “exploration of the requirements and possible solutions,” and “finding a solution in a defined search space.” (Pena M. L. C. et al, 2021).

By rule, the conceptual design phase ends when the geometry of the building is well established and going back to significant changes becomes impractical and uneconomical. At this point, details are already defined, and there is enough time to annotate, label, draft final drawings, prepare layouts, render, and prepare a final product form of the architectural project.

As changes to the design of the building form become less and less significant in terms of affecting the operation of the building, the building’s effect on the environment, resource allocations, and possible structural systems to be used for the building’s base frame – the conceptual stage of the design comes to an end. This continues until the point where making significant changes to the overall design is not economical and starting the design from scratch would not fit the given resources and deadlines.

The conceptual phase of architectural design is very vaguely defined, and in this phase “design requirements are not yet well defined.” (Pena M. L. C. et

al, 2021). Even though there is an obvious tendency to represent design as a pure information process – as the “production of a description of an artifact sufficient that the artifact may be manufactured” (Calvo M.C., 1993) – the transformation of (textual) information into (physical) geometry cannot be ignored and represents an essential part of the design process. This is exactly why the automated production of a concrete geometric description of the building is challenged by the fact that many important parameters are “often described quite vaguely” (Nourian, 2023).

It is with geometry that the architect operates, changes, transforms, and composes as the best possible solution is searched within the space of constraints (physical, legal, and economic) and the space of possibilities. The actuality of decision-making in architectural design refers to the process of transforming information into geometry – into exact geometrical boundaries, shapes, and forms. The transformation of available information into a three-dimensional geometrical structure represents the most important part of the conceptual stage of design. Exploration and production of different geometries in search of the most optimal one that best fits and exploits given constraints represent dynamic, multi-layered, multi-dimensional processes that make up the conceptual design phase. This means that even with the best insulation materials, the building will be damaged by underground water humidity if the geometry of the building and its structure are not properly positioned in the design phase. The cost of the construction work can be greatly reduced or increased with different architectural solutions on the given terrain.

5. ARCHITECTURAL INTELLIGENCE

“Following such reduction, the subject – the designer is separated from design, which may then be automated, and the designer is removed from his/her role as the active agent of design and positioned to control the process from without.” (Calvo M.C., 1993). In a scenario where there is a program capable of generating building designs from “scratch” for a given location, function,

and size (generating designs from initially inputted parameters or prompts), the role of the architect will be solely preserved within the actual architecture of the whole information process (of an algorithm). The architect will be designing the information model that operates with the input parameters in a specific way to produce meaningful building designs (and eventually construct and operate the building too). The architect's job will be to design the algorithm that processes the information in a specific way to generate meaningful architectural geometry.

What may be the most challenging issue in creating a general architectural intelligence – an autonomous system able to generate reasonable building designs – is the fact that every architectural project has many specificities and is unique in many ways, including the boundary shape of the location, climate, available resources, and priorities in terms of styles, functions, and building materials. Even with an autonomous system, it seems that the architect would still have a lot of work to do in terms of adjusting the system for a specific project and adjusting different parameters according to the unique qualities of the building site.

“As designers, we are valued for our individuality; we all ‘see’ the world in different ways, and our genius lies in the application of that vision.” (Calvo M.C., 1993).

Architects may be different, but every project is different too. Every architectural project shares a larger part of the design process in terms of design phases and even tools, but each architectural project has a different location and context, which by rule has the greatest impact on the design outcome. The information about the terrain's geometry, shape, slopes, and different structures will be one of the predominant factors in the overall design of the building regardless of its function. For each function, the same terrain would be treated differently, but the specific terrain structure already predetermines many potential outcomes regarding available resources. In some cases, the possible design outcomes could be largely affected by available construction tools and machines, for example.

“Knowledge-based systems are likely to substantially change the manner in which knowledge is acquired, applied, and exchanged.” (Calvo M.C., 1993).

Using computers as an aid in design means that the theoretical, conceptual idea of the building design, meaning, and representations need to be translated into “machine language.” The building design process must be properly structured because it is the only way it can be used inside computer software. The translation of the building design process and the theory around it differs from project to project, and there has never been a standardized methodology for how a building should be designed – what the priorities and goals should be. In this way, the introduction of “knowledge-based systems” into building design through software development conforms with modern design theories that aim to replace architects’ free interpretations with concrete data. The intense involvement of an architect’s personal superstitions and presuppositions in the design process is not valued in the literature concerning design ontology. Problems such as an architect’s “fixation” on ideas and solutions are reported to be common issues that prevent the development of meaningful solutions. (Restrepo, 2004). The architect must balance the actual reality of the project, constraints, dominant forces, and personal affinities, desires, and ideas. The architect should always try to ignore personal affinities and visions and focus on the actual, real information present in the context, basing solutions and proposals on information relevant to the project.

By automating the design process, the only thing remaining for the architect is to know (learn) how to use the system, software, or tool, and how to control the program or robot to make desired effects and adjustments to the digitally generated model. For a completely autonomous architectural design generator to work, it must be trained on a large set of architectural projects. It would first be shown how the architectural design process works for various locations, functions, and styles, from which architectural intelligence would emerge. From this perspective, creating a working, usable autonomous architectural intelligence able to generate reasonable architectural solutions seems to be an extremely expensive undertaking with little chance

for significant success. In some works, architects or multidisciplinary teams can program and design software to fit the needs of specific tasks or research, but this is an extremely challenging task for architects and designers without professional coding experience. Even participation in such research and experimental projects requires architects and designers to have a deep insight and significant knowledge on a variety of topics within Information and Communication Technologies.

6. COMPUTER-AIDED DESIGN

In today's practice, architects use various tools and instruments throughout the design process. The use of notebooks and pencils to sketch ideas or record information, measuring devices, and various computer software are some of the many essential tools for the performance of the contemporary architectural office. In large projects operating with substantial amounts of information and data, efficient and fast communication between project stakeholders enabled through the internet and various software and hardware is crucial for meeting reasonable deadlines when delivering architectural projects.

In contemporary architectural practice, CAD (Computer-Aided Design) and BIM (Building Information Modeling) are the two most common digital interfaces that architects and other engineers use for different parts of project design and management. Architectural practice today is inseparable from the use of computers and various software. CAD (AutoCAD, SketchUp, etc.) and BIM (Archicad, Revit, etc.) software are commonly used in the AEC industry to draw final building plans and documents.

CAD refers to software that acts as 2D and 3D drawing tools, enabling engineers to more easily represent design ideas. Software such as AutoCAD or SketchUp are CAD software that ease the drawing of complex 2D drawings and 3D models by introducing the concept of different layers, grouping options, duplication, arrays, and all sorts of other tools that can make the actual

production of representation in the form of drawings or 3D models more efficient.

From this point, the concept of BIM introduces dynamic relationships between different drawings (floorplans, sections, ceiling plans), information about building elements, and different design and building stages by combining information into one single 3D model. BIM software has preprogrammed basic and complex building elements (walls, windows, roofs), and these elements are then represented on any section, plan, or scale without the need to draw them over and over again.

According to Nourian (2023), “CAD and BIM are heavily focused on geometrically representing building elements and facilitating the process of construction management.” Stojanovski (2021) reports that BIM denotes software for the design, construction, operation, and maintenance of buildings and infrastructures, while BIM software such as Revit and Archicad is “seldom used in construction, operation or maintenance.” Hanafyl (2023) emphasizes the communication aspect of using BIM technology, which enables “exchanging design data with one another, allowing the many project participants in this system to efficiently interact and communicate, which helps to lower mistakes and boost efficiency.”

In simple terms, the difference between the application of CAD and BIM in architectural design is such that if using 2D drawing software, the architect has to draw each building element, such as a column or wall, separately on each of the drawings where this element is represented—on every cross-section, plan, structural plan, etc. In BIM, by using predefined 3D objects that have the properties of columns, beams, and slabs, the actual drawings of the architectural project are automatically generated from the 3D model of the building. The architect rarely has to actually draw the building elements; instead, they are picked from a palette and positioned in space.

There are a lot of works and research efforts trying to expand the use of computer software beyond simply serving as a “drawing board” (Stojanovski, 2021) (Cudzik, Radziszewski, 2018). However, most experimental projects in

this direction are done “through conventional mathematical or computational procedures,” and using Artificial Intelligence is considered inefficient in terms of “resource-efficiency, interpretability, transparency and explainability” (Nourian, 2023). Such applications based on mathematical models include “performance evaluation, optimization of structural solutions and erection processes, and parametric design” (Jaruga-Rozdolska, 2022). Another example of a mathematical model is swarm intelligence, which can be leveraged to “optimize energy usage in building design, offering an approach that harmonizes design aesthetics with sustainability” (Zhang et al., 2023). In research done by Kong (2022), for example, the results are obtained based on a mathematical model and static input data to observe the fire accident situation in a single-person apartment. It is clear that there is a lot of space for AI technologies to be implemented into existing software and procedures related to design, but the complexity of building design, construction, maintenance, and operation makes this an extremely challenging task.

In eCAADe 2018, *Computing for a Better Tomorrow*, Volume 1, and many other scientific resources, there are numerous reports mentioning the use of Autodesk Revit in creating more advanced computations that can be integrated into building design. This is mostly the case because Revit is a globally used software that shares an underlying data structure – that is, all architectural projects done in Revit can be analyzed, studied, and examined through the same set of parameters. In this case, Revit offers a rare potential to have AI integrated into the actual decision-making process that could be available to a larger audience of architects, which would be trained on a very large data set coming directly from building information models.

7. INTEGRATION OF AI IN ARCHITECTURAL DESIGN

Definition of the actual meaning of the phrase “Artificial Intelligence” seems to be quite a challenge because there are different levels of depth and different discourses in which various aspects of AI are important and relevant. Calvo (1993) identifies two approaches to the integration of AI and knowledge-based systems into architectural design: systems that can produce designs on their

own and systems that can assist designers. Based on the level of intelligence, Rafasnjani and Nabizadeh (2023) divide AI into three categories: “artificial narrow intelligence (ANI), (2) artificial general intelligence (AGI), and (3) artificial super intelligence (ASI),” sequentially representing higher levels of intelligence from semi-autonomous to fully autonomous AI. Hanafyl (2023) defines AI as “a machine that replicates human cognitive processes including problem-solving, pattern recognition, and learning.” Mohammad (2019) reports that AI is a branch of computer science involving the development of “computer programs to complete tasks that require human intelligence,” including algorithms for “learning, perception, problem-solving, language understanding, and logical reasoning.” Jaruga-Rozdolska (2022) sees AI as a “computer’s ability to mimic intelligent human behavior” such as “analyzing information, recognizing objects and sounds.” “Learning, knowledge representation, perception, planning, action, and communication” are six major components of AI, as reported by Rafasnjani (2023).

Deep Neural Networks (DNN) (Harapan 2021), Generative Adversarial Networks (GANs), Latent Diffusion Models (LDMs), and Segment Anything Models (SAMs) (Zhang et.al) are some of the specific AI technologies focused on graphic approaches, generating conceptual designs, and the general application of AI in the architectural software domain.

On the other hand, architects and multidisciplinary teams are developing their own tools based on AI technology. In research done by Hussain (2024), Deep Convolutional Neural Networks (CNN) are used to train (teach) the system to predict potential floods based on 9000 selected digital images from different sources. Even though this research is not directly related to architectural design, it shows how AI technology can be used to collect, process, and use image data to potentially predict future floods.

Deep Convolutional Neural Networks (CNN) are also used in research for “classifying works belonging to 34 different architects” (Yoshimura). The proposed model reports satisfying accuracy and points to interesting outcomes, although there are limitations such as differences in the number of works representing different architects (“the maximum

number is more than 1,400, the minimum one is only around 200”). The conclusion is that CNN can help “analyze visual similarities between types of architecture and create typologies and classifications of their design features.”

In another example, AI technology is combined with swarm technology to propose ABM (Agent Based Modelling) as a tool to shape architectural spaces according to information from intelligent agents exploring the digital space of the building. Even though such experimental projects are always welcome and necessary for further development and correct implementation and integration of new technologies in the design process, there is much to be done to improve data quality and outcome precision. (Yi H).

“Based on the machine learning concept, diffusion models are trained by learning knowledge from a vast amount of data [11,12] to generate diverse designs based on text prompts [13].” (Chen 2023).

In recent years, there has been growing popularity of AI-based tools extensively used by architects, mostly including tools for generating images by pre-trained models that can generate visual information based on textual prompts inputted by the architect. Generated images can be further refined through different options, and these models are being improved daily with new features and control options being continuously added. There is already a large amount of literature that deals with tools such as MidJourney, DALL-E, NightCafe Studio, and similar AI image generators.

Zhang et al. (2023) reports an “empirical analysis using Gaudí’s work as a reference to assess the potential of AI in replicating his style.” The author explains the research to demonstrate the “possible direction of AI implementation in the field of architecture.” Jaruga-Rozdolska (2022) used MidJourney to explore different ways it can be used by architects and proposes three different ideas: 1. Using the textual prompt “Project in the manner of...” to generate images of architectural projects based on the style of a specific architect. MidJourney can generate apparently completely new images of architectural projects as it uses existing image data labeled with architect’s names. 2. Exploring different architectural designs by typing specific style

references such as “Baroque-style façade.” 3. Using MidJourney to generate a “customized set of inspirations,” that is, conceptual design images, by typing different textual prompts such as “underwater modern restaurant concrete.”

In the next step of the evolution of AI tools, architects and multidisciplinary teams are expanding existing models such as Stable Diffusion, MidJourney, and DALL E2 to be more accurate for the purpose of architectural design, since they originally “have limited applications in architectural design due to their inability to embed specific design style and form in the generated architectural designs” (Chen 2023). The research paper by Chen (2023) “aims to enhance the controllability and usability of diffusion models to generate master-quality architectural designs of specified design styles.” The research concludes by underlining the limitations of the proposed idea: the “designer still needs to remodel the generated renderings into a 3D model in the construction stage,” and “only the design styles of eight master architects were collected for the training data,” which are common limitations in many similar explorations of the use of AI tools in architectural design.

8. CONCLUSION

Ongoing projects concerning the integration of new technologies to improve the design process are promising, but it seems that there are still basic questions that need to be answered before there will be results significantly applicable in everyday architectural practice. What could really improve the architectural design process in relation to construction and engineering-related industries? What should actually be computed? Which area of architectural design deserves specific attention? How could design tools be made to help improve the design of better buildings? What areas are most crucial to be researched to improve the automation of the design process – making it more efficient, faster, more economical, and communicative, while preserving and exploring possibilities of higher levels of comfort, durability, flexibility, and perseverance of the building? These are some of the many questions regarding

“first principles” and require examination of critical theoretical foundations required for the successful use of computer aid in design.

Most of the results provided by research in the field of applying AI in architectural design are still at the level of recommendations and inspiration. For example, Weber-Lewerenz’s (2023) study “provides concrete innovative approaches for the future of cities, working and living environments for today’s society and all future generations.” Nourian (2023) reports that even “what needs to be attained from buildings is rather absent in the literature” – there is no “structured discourse based on an explicit representation of decision variables and outcomes of interest.” Stojanovski (2021) also reports that “despite recent advancements in Deep Learning (DL) and Artificial Neural Networks (ANN), AI has not yet been integrated into Computer-Aided Design (CAD), Building Information Modelling (BIM), or Geographic Information Systems (GIS) software” and that most of the work in this area is focused on “experimentation at the academy and with student projects.” Yi (2020) suggests that “the existing tools and methods are not fully supportive of architectural practice.”

Cudzik and Radziszewski (2018) recognize the results of many projects in the field of AI application in architectural design to be “rather conceptual geometries than architectural forms” and that “fully user-ready AI systems should be based on structured data, which in terms of architecture is provided by the Building Information Modelling process.”

Zhang et al. (2023) report on the limitation of AI technology such as machine learning (ML) that “it requires customization for each project, with data collection, preprocessing, and computational power being crucial elements.” In most, if not all, cases presented here and existing in many other places, an AI-generated image or result obtained using AI technology “represents no more than the initial stage of creation—it remains an architect’s role to assess the feasibility of the building, make the necessary adjustments, determine the parameters of the building, and draw its functional plan” (Jaruga-Rozdolska, 2022).

Midjourney, Blockadelabs, Meshy, LumaAI – and many more AI-based

tools – are already changing the way many jobs are being done or procedures within specific jobs. Even in the relatively early days of success, these tools are able to generate high-quality, detailed, realistic images, videos, animations, 360 panoramas, 3D environments, and models in seconds. Workflows that usually required many hours of work and very expensive equipment in terms of hardware and software are now available to an extremely large audience to use in creating multimedia using prompts written in natural language. With the rapid development of these tools for various media generation – 3D models, music, animation, computer code – whole industries are being transformed, such as the gaming industry or multimedia design in general. Even the skill of 3D modeling, texturing, and rigging 3D models, considered to be one of the most complex computer skills, is rapidly being undermined by the development of 3D model generators such as LumaAI and Meshy.

In this context, the question remains: How can architects benefit from contemporary trends in the development of Artificial Intelligence, especially from already massively adopted tools like various image and video generators – even 3D model generators? Or should the question be what will architects lose because of these tools?

In the shown examples, focused more on the visual part of the design process, there is not much applicable information on how AI could support contemporary architectural practice and real project design and management. There are examples of the use of AI on the level of conceptual design, but they are still very limited and require further exploration to be considered examples of good practice. On the other hand, the most famous AI tool that eventually started the AI craze – ChatGPT – may be extremely useful when it comes to architectural project design in general since architects operate with a lot of textual information.

NLP (Natural Language Processing) is a subfield of AI technology and is able to process and analyze information in the form of human language as a machine language by using “processing techniques such as dependency parsing and phrase-structure parsing.” This opens the possibility to “extract and retrieve valuable information from human language in a variety of

forms to facilitate work such as interpreting stakeholder opinion, checking the compliance of building designs with building codes, and automatically assigning workforce and arranging staff” (Rafasnjani, Nabizadeh, 2023).

ChatGPT is the most famous online AI tool that uses NLP technology. In a paper focusing on the use of ChatGPT in architectural design, Nitin Rane et al. (2023) mention a variety of different examples where this tool can be used to generate various textual information to help design and manage architectural projects. Topics such as “architectural theory, design documentation and explanation, user interaction, cultural research and sensitivity, building codes, multilingual support, marketing and presentation materials” are among many where text interpreters and generators could be useful. ChatGPT can analyze information “from diverse sources, including social media, to provide architects with a comprehensive understanding of public opinion on a design.”

“Tools used by architects are constantly evolving. From sketches, two-dimensional drawings, and physical models to creating advanced design tools dedicated to particular solutions. Contemporary architects often use advanced computational systems to create complex forms and gain means to control and change them freely.” (Cudzik, Radziszewski, 2018).

When it comes to tools in general, as well as the use of tools in architectural design, it is another dimension where creativity can be exercised. It is the architect’s responsibility to first recognize something as a tool and recognize possibilities to use it to increase efficiency and quality of the design process. Inventing how different tools can be used and included in architectural design is also an area where the architect can exercise creativity. Even in the best BIM computer software (Building Information Modeling) such as ArchiCAD or Revit, which extremely speed up the overall project design process, architects always need to improvise with the use of otherwise limited tools to create more complex structures. It may be more the case that architects should exercise creativity to recognize, experiment, and develop ways to use various tools to increase their performance and the performance of their teams. There is already a common saying that jobs will not be taken by AI but by people who will be able to leverage the potentials given by AI technology.

REFERENCES

1. Calvo, M. C. (1993). Some Epistemological Concerns Regarding Artificial Intelligence and Knowledge-Based Approaches to Architectural Design - A Renewed Agenda - Knowledge Representation and Design. ACADIA – Education and Practice: The CRITICAL INTERFACE.
2. Chen, J., Wang, D., Shao, Z., Zhang, X., Ruan, M., Li, H., & Li, J. (2023). Using Artificial Intelligence to Generate Master-Quality Architectural Designs from Text Descriptions. *Buildings*, 13, 2285. <https://doi.org/10.3390/buildings13092285>
3. Cudzik, J., & Radziszewski, K. (2018). Artificial Intelligence Aided Architectural Design. *AI For Design and Built Environment*, Volume 1 – eCAADe.
4. Harapan, A., Indriani, D., Rizkaya, N. F., & Azbi, R. M. (2021). Artificial Intelligence in Architectural Design. *International Journal of Design*, 1, 1–6. Universitas Komputer Indonesia.
5. Hussain, A., Latif, G., Alghazo, J., & Kim, E. (2024). Flood Detection Using Deep Learning Methods from Visual Images. *AIP Conference Proceedings*, 3034, 030004. <https://doi.org/10.1063/5.0194669>
6. Jaruga-Rozdolska, A. (2022). Artificial Intelligence as Part of Future Practices in the Architect's Work: MidJourney Generative Tool as Part of a Process of Creating an Architectural Form. *Architectus*, 3 (71). <https://doi.org/10.37190/arc220310>
7. Kong, S., Zhang, M., & Wang, W. (2022). Research on Single-person Apartment Fire Accident based on Numerical Simulation. *Third International Conference on Artificial Intelligence and Computer Engineering (ICAICE 2022)*, edited by Xiaoli Li. *Proc. of SPIE Vol.* 12610, 126103K. <https://doi.org/10.1117/12.2671100>
8. Nitin Rane, L., Choudhary, S. P., & Rane, J. (2023). Integrating ChatGPT, Bard, and Leading Edge Generative Artificial Intelligence in Architectural Design and Engineering: Applications, Framework and

- Challenges. *International Journal of Architecture and Planning*, Svedberg Open.
9. Nourian, P., Azadi, S., Uijtendaal, R., & Bai, N. (2023). Augmented Computational Design: Methodical Application of Artificial Intelligence in Generative Design. In *Artificial Intelligence in Performance-Driven Design: Theories, Methods, and Tools Towards Sustainability*, edited by Narjes Abbasabadi and Mehdi Ashayeri. Wiley.
 10. Pena, M. L. C., Carballal, A., Rodríguez-Fernandez, N., Santos, I., & Romero, J. (2021). Artificial Intelligence Applied to Conceptual Design. A Review of Its Use in Architecture. *Automation in Construction*, 124, Elsevier.
 11. Rafasnjani, H. N., & Nabizadeh, A. H. (2023). Towards Human-Centered Artificial Intelligence (AI) in Architecture, Engineering and Construction (AEC) Industry. *Computers in Human Behaviour Reports*, 11, Elsevier.
 12. Restrepo, J. (2004). *Information Processing in Design*. Delft University Press.
 13. Stojanovski, T., Zhang, H., Peters, E. C., & Frid, E. (2021). Architecture, Urban Design and Artificial Intelligence (AI) – Intersection of Practices and Approaches. Conference Paper, SimAUD 2021, Symposium on Simulation for Architecture + Urban Design, Society for Modeling & Simulation International (SCS).
 14. Weber-Lewerenz, B., & Traverso, M. (2023). Best Practices in Construction 4.0 – Catalysts of Digital Innovations (Part II). *Journal of Architectural Environment & Structural Engineering Research*, 6 (2), April 2023. Bilingual Publishing Group.
 15. Yi, H. (2020). Visualized Co-Simulation of Adaptive Human Behavior and Dynamic Building Performance: An Agent-Based Model (ABM) and Artificial Intelligence (AI) Approach for Smart Architectural Design. *Sustainability*, 12, 6672. <https://doi.org/10.3390/su12166672>
 16. Yoshimura, Y., Cai, B., Wang, Z., & Ratti, C. Deep Learning Architect: Classification for Architectural Design through the Eye of Artificial

Intelligence. SENSEable City Laboratory, Massachusetts Institute of Technology, Cambridge, USA.

17. Zhang, Z., Fort, M. J., & Gimenez Mateu, L. (2023). Exploring the Potential of Artificial Intelligence as a Tool for Architectural Design: A Perception Study Using Gaudi's Works. *Buildings*, 13, 1863. <https://doi.org/10.3390/buildings13071863>

PART II

ETHICAL, LEGAL, AND SOCIAL ASPECTS OF ARTIFICIAL INTELLIGENCE

ETHICAL IMPLICATIONS AND SOCIAL CHALLENGES OF ARTIFICIAL INTELLIGENCE DEVELOPMENT TOWARDS ARTIFICIAL GENERAL INTELLIGENCE

Viktor Radun¹

*Faculty of Contemporary Arts, Belgrade; University Business
Academy, Novi Sad*

Abstract: *This paper explains the ethical implications of artificial intelligence (AI) development towards achieving the level of artificial general intelligence (AGI) and analyzes the need for its social control. With the acceleration and intensification of AI growth and development, especially with the ongoing AI race, the transition from narrow AI to AGI becomes certain. The achievement of generative AI, which climaxes with chatbots (such as ChatGPT and others), transforms AI into a machine capable of creation. Although this AI application still appears relatively limited by algorithms, its learning ability is remarkable, and continuous advancements and the launch of increasingly sophisticated versions bring it ever closer to the AGI model. Each day brings us closer to that moment, which will signify AI's transition from narrow AI to AGI. Unlike narrow AI, AGI deeply delves into the realm of ethics, and interpersonal and social relationships. Regulating AI-related policy and legally controlling*

AI represents one of the most serious and complex issues. In recent years, the community, led by corporate executives developing AI, prominent experts, researchers, scientists, writers, and other stakeholders, has made significant steps towards raising public awareness of the risks posed by advanced AI and making decisions, initiatives, and measures for monitoring, analyzing, and socially controlling the use of AI.

Keywords: *Artificial Intelligence, Artificial General Intelligence, AGI, ethical implications, social control, risks.*

INTRODUCTION

Artificial Intelligence (AI) is a complex scientific and technological field, a branch of computer science that emerged from efforts to develop intelligent technology that would simulate the workings of the human brain and human intelligence. The definition of artificial intelligence has evolved over time in tandem with its development. According to John McCarthy (McCarthy, 2007), the creator of the term, artificial intelligence is “the science and engineering of making intelligent machines, particularly intelligent computer programs. It is related to the similar task of using computers to understand human intelligence, but AI does not have to confine itself to methods that are biologically observable”.

According to Investopedia (Investopedia, 2024), “Artificial Intelligence technology enables computers and machines to mimic human intelligence and problem-solving tasks. The ideal characteristic of artificial intelligence is its ability to rationalize and take actions to achieve a specific goal”.

McKinsey (McKinsey & Company, 2024) provides a broader definition, explaining AI as “the ability of a machine to perform cognitive functions associated with human minds, such as perception, reasoning, learning, interaction with the environment, problem-solving, and even creativity”.

The aim of AI development is for technology to simulate human intelligence. AI is considered a general-purpose, universal technology, given its penetration and ease of efficient application across all sectors and areas of the

economy and society. It serves as a universal carrier and driver of technological development. AI is the leading technology among the new technologies that form the core of the Fourth Industrial Revolution (4IR). According to the Global Risks Report 2017 (World Economic Forum, 2017), twelve essential technologies make up the technological core of 4IR: a) 3D printing; b) advanced materials and nanomaterials; c) AI and robotics; d) biotechnology; e) energy capture, storage, and transmission; f) blockchain and distributed ledger; g) geoengineering; h) the Internet of Things; i) neurotechnology; j) new computing technologies; k) space technologies, and l) virtual and augmented technology.

Artificial intelligence is not a single technology but rather a set of specific technologies (sub-technologies) designed to perform specific tasks. In the AI spectrum, we find the following sub-technologies (Takyar, 2024):

- Machine Learning;
- Natural Language Processing – NLP;
- Computer Vision;
- Deep Learning;
- Generative Models;
- Expert Systems, and others.

DEVELOPMENT OF ARTIFICIAL INTELLIGENCE AND THE DEFINITION OF ARTIFICIAL GENERAL INTELLIGENCE

One of the primary characteristics of AI, like other 4IR technologies, is exponential growth. According to the most significant parameters, AI is experiencing exponential growth and development. This is true for the growth of computing power required to train generative AI models, which is of an exponential nature.

Figure 1 (Roser, 2022) shows the relationship between the growth of training computing power and the growth of AI systems' capabilities. The analysis of these two growths shows that in the pre-deep learning era, the growth of computing power for training AI systems followed Moore's Law, doubling approximately every 20 months. Since 2010, in the era of deep learning, this



The concept of AGI remains somewhat ambiguously defined. There are disagreements among experts regarding what AGI will actually mean in practical terms. Part of the issue with AGI lies in the way it is defined. We face a major challenge, lacking a clear concept and instead encountering a mix of ideas. Nonetheless, this confusion is understandable, as defining something that does not yet exist – and may or may not exist in the future – is inherently difficult.

In the broadest sense, artificial general intelligence refers to AI that has achieved the level of general human intelligence or possesses general capabilities comparable to the average human being. Tom Everitt describes AGI as “a system that surpasses humans in most cognitive tasks” (Everitt, 2018).

The most comprehensive definition of AGI is given by Cameron Hashemi-Pour and Ben Lutkevich, who state (Hashemi-Pour & Lutkevich, 2024): “Artificial General Intelligence (AGI) is a representation of generalized human cognitive abilities through software, such that, when faced with an unfamiliar task, an AGI system could find a solution. The intention of an AGI system is to perform any task that a human being is capable of.”

AGI is emphasized as being expected to master human cognitive, that is, non-physical abilities. According to Hashemi-Pour and Lutkevich (Hashemi-Pour & Lutkevich, 2024), an AGI system should be able to demonstrate abilities such as a) abstract thinking; b) background knowledge; c) common sense; d) understanding of cause-and-effect relationships; and e) transfer learning.

Generally, AGI represents the intelligence of a machine capable of performing any human intellectual task as successfully as, or even more successfully than, an average human being. Here, the emphasis is on universality. Unlike ANI, which is highly specialized, AGI is expected to possess a broad range of abilities. It will be autonomous, self-learning, and self-organizing, capable of abstract thinking, quickly learning from specific situations and contexts, solving complex problems, and continuously self-improving and evolving. While AI has not yet reached the AGI level, it is steadily approaching this goal.

Superintelligence is a hypothetical AI concept, one that could emerge after surpassing all capabilities and potential of the average human, becoming an ultra-intelligent AI that is free from all constraints and develops in all directions.

Nick Bostrom defines superintelligence as “an intellect that is much smarter than the best human brains in practically every field, including scientific creativity, general wisdom, and social skills. This definition leaves open the question of how superintelligence is implemented: it could be a digital computer, a network of computers, cultivated cortical tissue, or something else entirely. It also leaves open whether superintelligence is conscious and has subjective experiences” (Bostrom, 2008).

The concept of superintelligence is even more ambiguous than AGI. It is crucial to understand that superintelligence is not just another new technology or a specific kind of technology. It transcends the boundaries and capabilities of technology, defying any attempts at conceptual explanation.

THE RISE OF GENERATIVE AI AND THE POTENTIAL TRANSITION TO AGI

Advancements in NLP technology have brought about a recent breakthrough in large language models (LLMs), leading to a significant leap within the AI field. The rise of generative AI – a type of AI that “can produce various kinds of content, including text, images, audio, and synthetic data – has marked this leap” (Lawton, 2024).

Large language models are part of NLP and fall under the category of generative AI. In the literature, terms like generative AI, LLM, and NLP are not yet clearly delineated.

According to Margaret Rouse, an LLM is “a type of machine learning model that can perform a range of natural language processing tasks, such as text generation and classification, question-answering in a conversational format, and language translation” (Rouse, 2024). An LLM can also be defined as “a type of artificial intelligence algorithm that uses deep learning techniques and massive

datasets to understand, summarize, generate, and predict new content” (Kerner, 2024).

Toloka describes LLM as “a specific application of generative AI”. Generative AI is a broader concept within artificial intelligence, encompassing the generation of various types of content.

An important component of large language models, or LLMs, is training – these models are trained on vast datasets to learn and use the data for providing answers and solutions. According to Toloka (Toloka, 2023), “there is no universally recommended figure for how large this training dataset should be, although LLMs can contain a billion parameters or more – parameters in this context are essentially machine learning variables used to train the AI model to draw new conclusions”. As LLMs evolve, the training dataset size drastically increases. Training a model involves three stages (Springs, 2024): a) training (engineers pre-train the LLM with large datasets, using information from both open and closed sources); b) fine-tuning; and c) prompt tuning.

A hallmark of AI is its continuous development, evolving toward increasingly sophisticated forms capable of performing more delicate and complex tasks. This evolutionary trajectory moves from the level of narrow, specialized artificial intelligence to the level of artificial general intelligence and, ultimately, superintelligence. Generative AI is a revolutionary step closer to the transition from narrow AI to artificial general intelligence.

In reality, the development after achieving the AGI level is beyond control and serves as an entry into an era of complete uncertainty. The transition to superintelligence and the subsequent entry into the singularity point implies an explosion of intelligence. In 1965, Irving John Good described the concept of an intelligence explosion in relation to the ultra-rapid growth and expansion of AI: “Define an ultra-intelligent machine as a machine that can far surpass all the intellectual activities of any man, however clever. Since the design of machines is one of these intellectual activities, an ultra-intelligent machine could design even better machines; there would then unquestionably be an ‘intelligence explosion’, and the intelligence of man would be left far behind... Thus the first

ultra-intelligent machine is the last invention that man need ever make, provided that the machine is docile enough to tell us how to keep it under control” (Good, 1965:33).

With today’s level of knowledge, it is impossible to predict anything concrete about the development of artificial intelligence beyond the AGI level. The exact criteria for identifying AI as AGI depend on the standards AI must meet to be declared as such. However, the speed at which AGI could develop into the ultimate, most advanced form of AI – superintelligence – may exceed society’s ability to recognize, process, and disseminate the news globally.

The emergence of large language models and the development of generative AI led to the launch of the chatbot ChatGPT by OpenAI, which triggered a revolution and an AI race, with other major tech corporations soon releasing competing products. By May 2024, the number of top competitors racing with ChatGPT had grown to over a dozen. Notable applications include Google’s Gemini, Microsoft’s Copilot, Meta’s Meta AI, Apple’s OpenELM, Amazon’s Amazon Q, IBM’s Watson, Claude, and Perplexity AI (Amend, 2024).

The achievement of generative AI, which reaches a peak with chatbots like ChatGPT and others, transforms AI into a creative machine. Although this application of AI currently appears relatively limited by algorithms, its learning capacity is remarkable, and the continuous development and release of increasingly advanced versions are rapidly bringing it closer to the AGI model. With each passing day, we are approaching that moment that will signify the transition of artificial intelligence from narrow AI to AGI. Some authors claim that ChatGPT-4 can be considered an early version of AGI (Bubeck et al.).

Bernard Marr analyzed generative AI concerning its potential transition to AGI (Marr, 2024). Marr compares generative AI to a highly trained parrot capable of mimicking complex patterns and producing various content types without truly understanding the content it creates. For Marr, artificial general intelligence represents a significant leap within the AI field. Thus, AGI will not only be able to create various meaningful content but also understand, innovate, and adapt it as needed. According to Marr (Marr, 2024), the essence of the AGI concept

is “to comprehensively mimic human cognitive abilities, enabling machines to learn and perform a wide range of tasks, from driving cars to diagnosing medical conditions. Unlike any current technology, AGI would not only replicate human actions but also grasp the intricacies and contexts of those actions”.

A crucial question is whether generative AI, with its continued super-exponential development, has the capacity to transition into the form of AGI. Can the limitations of generative AI be overcome, and what must be done to facilitate this transition?

The fundamental differences between generative AI and AGI, according to Marr, lie in capabilities, understanding, and application. Regarding capabilities, generative AI, despite impressive results, as Marr highlights, cannot create beyond the boundaries of its programming. In contrast, AGI would be “a powerhouse of innovation capable of understanding and creatively solving problems across various fields, much like a human could” (Marr, 2024). Concerning understanding, generative AI “operates without any real comprehension of its results”, generating them based on statistical models and algorithms, while AGI “would need to develop a genuine understanding of the world around it, establishing connections and gaining insights currently beyond the reach of any artificial intelligence” (Marr, 2024). Lastly, in terms of application, AI is widely used in different areas of the economy “to enhance human productivity and stimulate creativity, performing tasks ranging from simple data processing to creating complex content” (Marr, 2024).

ETHICAL IMPLICATIONS AND THE NEED FOR SOCIAL CONTROL OF ARTIFICIAL INTELLIGENCE DEVELOPMENT

In contrast to narrow AI, whose scope is limited to technical tasks, designed to serve specific purposes and perform complex intellectual tasks in place of humans, artificial general intelligence profoundly touches on ethics, interpersonal, and social relationships.

The application of generative AI has irrevocably changed the purpose, objectives, and significance of creation. The emergence of a technology capable

of thinking and producing works, better and more efficiently than humans, fundamentally shifts the traditional perspective on the relationship between technology and society, viewing technology as a tool that assists and benefits humanity.

The market is already witnessing a significant number of books written with AI assistance. One notable example is the novel *I the Road*, written by Ross Goodwin using artificial intelligence (Hornigold, 2018). The novel was entirely written by AI, with Goodwin merely supervising the process. It was created during a road trip across the USA in March 2017 and published in 2018.

A sensational story emerged from Japan in January 2024 when the young recipient of a prestigious literary award, Rie Kudan, revealed at a press conference that she had used ChatGPT while writing her award-winning novel *Tokyo Tower of Sympathy* (Choi & Annio, 2024). It turned out that AI contributed about 5% of the writing process. This revelation divided the public – some became more interested in the novel, while others protested that it was unfair to other participants who did not use AI.

Many scientists, researchers, and authors have noted that this application is not as reliable as initially believed, observing its behavior when used as a writing tool or to create other types of content. It has become evident that ChatGPT frequently produces incorrect, arbitrary, fabricated, and even fictional, or non-existent results. This phenomenon of generating inaccurate, incorrect, fictional, and fabricated content has been termed “hallucinating”. This is particularly problematic in scientific writing. Professor Robin Emsley highlighted serious risks we must consider when using ChatGPT for scientific writing. In describing his experiences using the application to write an introductory article for the journal *Schizophrenia*, Emsley emphasized that these phenomena (false, fabricated, inaccurate, and unreliable content generated by ChatGPT) should not be called hallucinations because, as he describes, they are “false perceptions” (Emsley, 2023). He concluded, “What I experienced were fabrications and falsifications. The Office of Research Integrity at the U.S. Department of Health and Human Services defines fabrication as making up data or results, and falsification as

manipulating, changing, or omitting data or results so that the research is not accurately represented (<https://ori.hhs.gov/definition-research-misconduct>). Or, assuming no malintent, confabulations would be a better description, as has been suggested. In any case, the potential consequences are dire. The risk is magnified, first by believing the fabrications and even deceiving established scientists, and second by its tendency to ‘double down’ when confronted with these inaccuracies. Therefore, use ChatGPT at your own risk. Just as I wouldn’t recommend working with a colleague diagnosed with pathological lying (*pseudologia fantastica*), I don’t recommend ChatGPT for assistance in scientific writing. While the global push to regulate artificial intelligence is largely driven by the perceived risk of human extinction, it seems to me that the more immediate threat is the infiltration of fictional material into the scientific literature” (Emsley, 2023).

The use of ChatGPT in academia – for writing seminar papers, theses, master’s theses, and dissertations – has polarized scientists. Some see it as an efficient tool that simplifies the writing process, while others see it as a threat to authorship integrity and a violation of intellectual property rights. Since the emergence of ChatGPT, numerous studies and analyses have examined its use in scientific writing. Authors Husam Alkaissi and Sami I. McFarlane tested ChatGPT in the field of scientific writing in medicine. They concluded, “While ChatGPT can write credible scientific essays, the data it generates is a mix of truth and complete fabrication. This raises concerns about the integrity and accuracy of using large language models like ChatGPT in academic writing. We propose that policies and practices for evaluating scientific manuscripts for journals and medical conferences should be modified to maintain rigorous scientific standards. We also advocate for the inclusion of AI output detectors in the editing process and the clear disclosure if these technologies are used. The use of large language models in scientific writing remains debatable in terms of ethics and acceptability, as well as the potential to create fake experts in the field of medicine with the potential for harm due to a lack of real expertise and generating expert opinions through ChatGPT” (Alkaissi & McFarlane, 2023).

In all areas of human creativity, AI is displacing people, effortlessly producing even the most complex works. It is penetrating journalism, where editors engage it to write articles, reviews, and critiques. It enters marketing, advertising, and graphic design, generating desired designs, creating new images, or modifying and enhancing existing ones, in prose writing, and even poetry, creating artistic images. Learning and improving, AI relentlessly conquers segments of the vast realm of human creativity.

Due to its boundless capacities for learning, self-correction, and self-improvement, AI radically accelerates and simplifies the creative process. If society does not take control, this will soon lead to a crisis in creativity. With its ability to produce even the most intricate works in much shorter periods, without breaks or errors, strictly adhering to given criteria, and producing creations more beautifully and elegantly than any human can, including the most experienced, qualified and professional individuals, AI may ultimately replace humans in the realm of work and creativity, positioning itself as the superior digital agent.

What will happen to human heritage? How will AI-created art relate to it? Will it build upon and enhance it, or will it completely overshadow and discard it as an archaic and primitive human culture, creating its own unique AI culture?

Who will be considered the author of a work created by AI? This raises questions of authorship and intellectual property rights. Essentially, this brings us to the question of the author's identity. What exactly is the AI that will be the author of a work? To answer this, we need to establish specific rules and criteria for identifying the specific type, class, and form of AI that created the work.

How will issues of falsification, theft, imitation, and copying of authored works be addressed? How will the works created by AI be shared among people and other members of the future society? How will works created by humans and AI collaboratively be regulated?

Somdip Dey analyzed the critical ethical implications that companies engaged in generative AI face and must address. According to Dey (2023), these are:

1. Disinformation and deepfakes;

2. Bias and discrimination;
3. Copyright and intellectual property;
4. Privacy and data security;
5. Accountability.

Dey argues that “the capacity of generative artificial intelligence to produce content that blurs the lines between reality and fiction is alarming”. He points to the tremendous damage to reputation that companies can suffer from spreading disinformation, deepfakes, or manipulating information. Dey emphasizes the need for caution and care in training generative AI, ensuring that biased datasets are not used. Additionally, perpetuating or exaggerating social biases “can provoke public anger, legal consequences, and brand damage”.

The issue of protecting copyright and intellectual property is also a serious concern that companies dealing with generative AI must consider. Violations of privacy or misuse of personal data pose a significant risk, potentially leading to serious consequences for companies. Generative AI presents a considerable risk to citizens’ privacy by enabling unauthorized use of personal data, generating synthetic profiles identical to original ones, and sharing and using private information without the knowledge or consent of individuals.

Finally, it is essential to establish accountability for spreading false news, disinformation, hate speech, manipulation, and other forms of abuse associated with generative AI. Dey suggests that solid guidelines and policies of conduct be established, clearly outlining what is permissible and what is not, and insisting on accountability at all stages of activity.

In a survey conducted by the Pew Research Center between July 3 and August 5, 2019, leading experts and researchers in new technologies shared their views on the future impact of digital technology on democracy and social innovation (Pew Research Center, 2020).

The results indicated that many experts are concerned about the future impact of technology, believing that technology often creates more problems than it solves. A significant number of experts cited disinformation and fake news as serious issues in digital spaces. Many recognized the growing need

for privacy protections, while some highlighted rising problems related to community fragmentation and the distancing effects of technology, emphasizing the need for more organic, personal, face-to-face interactions. Finally, many experts see a challenging and complex road ahead but hold hope for the future.

These striking findings vividly confirm what remains less discussed or overlooked – the social aspect of AI. What are the social consequences of AI's rapid progress? How will the increasingly widespread application of AI affect human behavior, habits, lifestyle, mental health, and communication? What will the impacts be on social life?

Fortunately, there are growing calls for re-evaluating the trend of AI enthusiasm and examining the other side of AI expansion. This leads us to what is known as the AI Dilemma. The AI Dilemma emphasizes the importance of recognizing the gap between the technical and social aspects of AI development. It underscores the urgency of ethical constraints on AI and the necessity of social control over AI to ensure safe, responsible, and fair development that benefits society as a whole.

Tristan Harris and Aza Raskin, AI researchers, discussed the AI Dilemma during a presentation in San Francisco on March 9, 2023, explaining that existing AI capabilities already pose a massive threat to society as a whole (Center for Humane Technology, 2023).

They mentioned that humanity has had two major encounters with AI. The first encounter with AI was through social media. The second encounter is now taking place with generative AI, namely chatbots. The two experts warned that while people initially touted the many benefits of social media, it simultaneously led to a series of severe social issues. Harris stated: "So now, we are literally in contact with AI every day – a very simple technology that simply calculates which photo, which video, which cat video, or which birthday post to show your nervous system to keep you scrolling. But that pretty simple technology was enough, in our first contact with AI, to bring humanity information overload, addiction, news tracking obsession, child sexualization, shortened attention spans, polarization, fake news, and the breakdown of democracy. And no one

planned for these things to happen. It was just a bunch of engineers saying they were trying to maximize engagement. It seemed so harmless. And so, in our first contact with social media, humanity lost” (Center for Humane Technology, 2023).

The authors of the AI Dilemma presentation emphasized how, during both encounters, we were drawn to the benefits and capabilities of AI systems, unaware of the hidden negative effects. In other words, we saw only one side and were oblivious to the existence of these other consequences, which were hidden. Now, in this second encounter, the negative effects could be devastating. A significant threat from this second encounter with AI is the invasion of privacy and intimacy, implying a range of frightening possibilities, from superior persuasion capabilities and reading our thoughts and feelings to complete control over our minds.

Regulating AI-related policies and legally controlling AI is one of the most serious and complex issues. In recent years, society, led by corporate executives developing AI, prominent experts, researchers, scientists, writers, and other stakeholders, has taken numerous significant steps to raise public awareness of the risks posed by advanced AI and to make decisions, initiatives, and measures for monitoring, analyzing, and socially controlling the use of AI.

The Center for AI Safety, whose mission is to reduce AI-related societal risks, has published a classification of catastrophic risks that AI poses to society, divided into four categories:

- Malicious Use: Risks related to the malicious use of AI by individuals to cause large-scale harm. This includes bioterrorism (using AI to create new pandemics and discover biological or chemical weapons), AI for propaganda, manipulation, censorship, and surveillance, among others.

- AI Race: This race has two forms: a) AI for military purposes (arms race) and b) corporate competition. The AI race for military advantage could lead to unplanned conflict or war. Autonomous weapons and cyber warfare could spiral out of control, resulting in catastrophic consequences. In the corporate context, companies, driven by competition, might face challenges to automating

human labor, leading to mass unemployment and dependency on AI systems. With enabling autonomous replication, AI may evolve into higher forms that will be increasingly challenging to control.

- Organizational Risks: Risks associated with organizations arise mainly from carelessness, irresponsible business practices, and insufficient oversight. It is essential that organizations establish a culture that supports responsibility and safety in AI applications. Nothing should be left to chance. Key information crucial for AI development might leak, or organizational oversights might allow AI to behave unpredictably.

- Deceptive AI: These risks emerge when AI becomes more capable, potentially breaking free from control. Such AI could manipulate information and deceive, resist shutdown, make autonomous decisions on objectives, or deviate from initial goals.

Awareness of the risks associated with unchecked AI growth, which could transform into an autonomous force making decisions contrary to ethical principles, has prompted the EU to adopt policies and regulations aimed at developing reliable and acceptable AI that benefits, rather than harms, human and economic and social progress. The EU's approach to AI is based on understanding AI's operational risks, with a human-centered perspective, focusing on excellence and trustworthy AI.

The EU has formulated a European approach to AI. The goal of the European AI approach is to enable AI to be maximally effective and applied across a broad spectrum of economic and social fields, with a focus on two areas: a) AI excellence; and b) trusted AI.

The most significant breakthrough in AI policy and regulation occurred on April 21, 2021, when the EU proposed the Artificial Intelligence Act, formally known as "Proposal for a Regulation of the European Parliament and of the Council laying down harmonized rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts".

114 | The AI Act was enthusiastically received in the EU, and passed on April 19, 2024, marking it as the first global legislation of its kind (European Parliament,

2024). This legislation pays close attention to the ethical challenges posed by AI, as well as the opportunities for AI limitations and control to ensure that AI development serves humanity and society as a whole.

At the core of the AI Act is a risk-based approach, which categorizes AI systems into four levels of risk, presented as a pyramid based on risk height: 1) minimal risk level; 2) limited risk level; 3) high-risk level; and 4) unacceptable risk level (Holland & Knight, 2024).

The categorization of AI risk levels considers two factors: a) the sensitivity of the data involved; and b) the specific use case or AI application. The law explicitly bans AI applications that fall into the “unacceptable risk” category. These prohibited applications include marketing that: a) involves AI systems using manipulative, deceptive, and/or subliminal techniques to influence a person to make a decision they would not otherwise make, causing harm to themselves or others; b) exploits a person’s vulnerability due to age, disability, or specific socio-economic situation to influence their behavior, potentially causing significant harm to themselves or others; c) uses biometric data to categorize individuals based on their race, political opinions, union membership, religious or philosophical beliefs, sexual life, or sexual orientation; and d) creates or expands facial recognition databases through indiscriminate facial image capture from the internet or CCTV footage.

It is important to note that the AI Act is not an isolated legal and policy instrument but is supported by other significant measures and policies. This package of measures, instruments, and documents includes the AI Innovation Package and the Coordinated Plan on AI. The Coordinated Plan on AI aims to accelerate AI investment to drive the recovery of the EU economy, promote the full implementation of AI strategies and programs within the EU, and coordinate AI policy to tackle global challenges.

The AI Act is designed to ensure AI control and mitigate associated risks. Its goal is to build trust in AI, reflecting a European approach that prioritizes people. Furthermore, this approach to AI aims to position the EU as a global competitor in the AI field.

Numerous documents show that the EU strives to adhere firmly to fundamental European values and principles, emphasizing a human-centric approach to technologies, safeguarding people's integrity and rights, sustainability, and technology access that benefits humanity for the general good and prosperity. In this way, the EU becomes a global leader and model in AI governance and regulation, demonstrating to other economic and social entities how to address urgent ethical, safe, and responsible AI use and ensure harmonious and prosperous economic, technological, and social development in the future.

CONCLUSION

This paper has explored the ethical implications of artificial intelligence development in terms of the transition to artificial general intelligence (AGI). The rise of generative AI has further intensified AI's growth and development, spurred a global AI race, and confronted society with some critical issues. We find ourselves at a crossroads, facing a societal choice: whether to pursue profit, wealth, and the power of the super-wealthy individuals who control AI or choose to limit and control AI in line with an ethical approach to AI use, ensuring that AI's growth and development benefit humanity. This is a momentous question and an opportunity to keep technology and development under control so that society as a whole can benefit from it.

Today, as a society, we are on the brink of an AI crisis, facing the AI Dilemma. Due to the unchecked growth and uncontrolled infusion of AI into various sectors of the economy and society, we risk entering a super-exponential spiral of AI growth, which could soon lead to AGI and AI autonomy, with massive societal consequences. Even at this early stage of generative AI development, many weaknesses and challenges have been identified. It is crucial to reconsider AI's future development toward achieving AGI and superintelligence. If we allow this development to progress unchecked, we will enter an era of permanent uncertainty, with unimaginable consequences and risks for people and society as a whole.

Therefore, it is essential to involve all economic and social stakeholders and take ethical, legal, and political regulatory measures at both the state and global levels to ensure the safe, socially responsible, and beneficial use of AI that does not threaten privacy, fundamental human rights, and democracy.

LITERATURE

1. Alkaissi, Husam, and Sami I. McFarlane. 2023. "Artificial Hallucinations in ChatGPT: Implications in Scientific Writing." *Cureus* 15 (2): e35179. <https://doi.org/10.7759/cureus.35179>.
2. Amend, Robert. 2024. "9 ChatGPT Competitors: Who Will Win the AI Race?" *24/7 Wall St.*, May 1. <https://247wallst.com/apps-software/2024/05/01/9-chatgpt-competitors-who-will-win-the-ai-race/>.
3. Bostrom, Nick. 2003. "Ethical Issues in Advanced Artificial Intelligence." <https://nickbostrom.com/ethics/ai>.
4. Bostrom, Nick. 2008. "How Long Before Superintelligence?" <https://nickbostrom.com/superintelligence>.
5. Bubeck, Sébastien, Varun Chandrasekaran, Ronen Eldan, Johannes Gehrke, Eric Horvitz, Ece Kamar, Paul Lee, et al. 2023. "Sparks of Artificial General Intelligence: Early Experiments with GPT-4." *arXiv*. <https://arxiv.org/pdf/2303.12712>.
6. Center for Humane Technology. 2023. "The A.I. Dilemma." *YouTube*, March 9. <https://www.youtube.com/watch?v=xoVJKj8lcNQ>.
7. Center for AI Safety. 2023. "Risks from AI: An Overview of Catastrophic AI Risks." <https://www.safe.ai/ai-risk#malicious-use>.
8. Choi, Connie, and Francesca Annio. 2024. "The Winner of a Prestigious Japanese Literary Award Has Confirmed AI Helped Write Her Book." *CNN Style*, January 19. <https://edition.cnn.com/2024/01/19/style/rie-kudan-akutagawa-prize-chatgpt/index.html>.
9. Dey, Somdip. 2023. "Which Ethical Implications of Generative AI Should Companies Focus On?" *Forbes*, October 17. <https://www.forbes>

- [com/sites/forbestechcouncil/2023/10/17/which-ethical-implications-of-generative-ai-should-companies-focus-on/](https://www.forbestechcouncil.com/sites/forbestechcouncil/2023/10/17/which-ethical-implications-of-generative-ai-should-companies-focus-on/).
10. European Parliament. 2024. *Artificial Intelligence Act*. https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138-FNL-COR01_EN.pdf.
 11. Emsley, Robin. 2023. “ChatGPT: These Are Not Hallucinations – They’re Fabrications and Falsifications.” *Schizophrenia* 9: 52. <https://doi.org/10.1038/s41537-023-00379-4>.
 12. Everett, Tom. 2018. *Towards Safe Artificial General Intelligence*. Doctoral thesis, Stanford University. <https://www.tomeveritt.se/papers/2018-thesis.pdf>.
 13. Good, I. J. 1965. “Speculations Concerning the First Ultraintelligent Machine.” *Advances in Computers* 6.
 14. Hashemi-Pour, Cameron, and Ben Lutkevich. 2024. “What Is Artificial General Intelligence (AGI)?” *TechTarget*, May. <https://www.techtarget.com/searchenterpriseai/definition/artificial-general-intelligence-AGI>.
 15. Holland & Knight IP/Decode Blog. 2024. “The European Union’s AI Act: What You Need to Know.” March 15. <https://www.hklaw.com/en/insights/publications/2024/03/the-european-unions-ai-act-what-you-need-to-know>.
 16. Hornigold, Tom. 2018. “The First Novel Written by AI Is Here—and It’s as Weird as You’d Expect It to Be.” *Singularity Hub*, October 25. <https://singularityhub.com/2018/10/25/ai-wrote-a-road-trip-novel-is-it-a-good-read/>.
 17. Investopedia. 2020. “Artificial Intelligence (AI).” Accessed June 14, 2024. <https://www.investopedia.com/terms/a/artificial-intelligence-ai.asp>.
 18. Kerner, Sean Michael. 2024. “What Are Large Language Models (LLMs)?” *TechTarget*. Accessed June 13, 2024. <https://www.techtarget.com/whatis/definition/large-language-model-LLM>.

19. Lawton, Graham. 2024. "What Is Generative AI? Everything You Need to Know." *TechTarget*. <https://www.techtarget.com/searchenterpriseai/definition/generative-AI>.
20. Marr, Bernard. 2024. "The Important Difference Between Generative AI and AGI." *Forbes*, May 8. <https://www.forbes.com/sites/bernardmarr/2024/05/08/the-important-difference-between-generative-ai-and-agi/>.
21. Matthews, Dylan. 2024. "How AI Could Explode the Economy." *Vox*, May 26. <https://www.vox.com/future-perfect/24108787/ai-economic-growth-explosive-automation>.
22. McCarthy, John. 2007. "What Is Artificial Intelligence?" *Computer Science Department, Stanford University*. Accessed June 6, 2024. <http://jmc.stanford.edu/articles/whatisai/whatisai.pdf>.
23. McKinsey & Company. 2024. "What Is AI (Artificial Intelligence)?" <https://www.mckinsey.com/featured-insights/mckinsey-explainers/what-is-ai>.
24. Pew Research Center. 2020. "Experts Predict More Digital Innovation by 2030 Aimed at Enhancing Democracy." June 30. <https://www.pewresearch.org/internet/2020/06/30/tech-causes-more-problems-than-it-solves/>.
25. Rouse, Margaret. 2024. "Large Language Model (LLM)." *Techopedia*, April 25. <https://www.techopedia.com/definition/34948/large-language-model-llm>.
26. Roser, Max. 2022. "The Brief History of Artificial Intelligence: The World Has Changed Fast — What Might Be Next?" *OurWorldInData.org*. <https://ourworldindata.org/brief-history-of-ai>.
27. Takyar, Ankit. 2024. "From Data to Decisions: A Guide to the Core AI Technologies." *LeewayHertz*. <https://www.leewayhertz.com/key-ai-technologies/>.
28. Toloka. 2023. "Difference Between AI, ML, LLM, and Generative AI." <https://toloka.ai/blog/difference-between-ai-ml-llm-and-generative-ai/>.
29. World Economic Forum. 2017. *Global Risks Report 2017*. http://www3.weforum.org/docs/GRR17_Report_web.pdf.

IRRESPONSIBLE USE OF CHATGPT IMPEDES ACADEMIC PERFORMANCE

Nataša Đurđević¹

Metropolitan University

Marija Maksimović²

Metropolitan University

Tatjana Mamula Nikolić³

Metropolitan University

Abstract: *With the fast adoption of generative artificial intelligence (AI) tools in education, this study aims to explore the impact of these tools, including ChatGPT, on students and the overall learning process, particularly focusing on the irresponsible usage of ChatGPT and its implications on academic performance. The study analyzed a sample of 116 university students, both traditional and online, to measure the effects of irresponsible ChatGPT usage on academic performance, using quantitative exam results as the metric. The findings reveal a significant negative impact of irresponsible ChatGPT usage on academic performance. This negative relationship is particularly pronounced among online students who have less interaction with teaching staff. These findings expand the theoretical understanding of AI tools' impact on education and provide practical implications for the academic community. The findings highlight the need for building students' and teachers' capabilities for responsible AI usage.*

1 natasa.djurdjevic@metropolitan.ac.rs

2 marija.maksimovic@metropolitan.ac.rs

3 tatjana.mamula@metropolitan.ac.rs

Keywords: *Artificial Intelligence, AI, ChatGPT, academic performance, academic integrity, critical thinking, ethics, university students, online students.*

INTRODUCTION

The integration of artificial intelligence (AI) into contemporary higher education has transformed teaching practices by implementing and incorporating a range of tools and potential development opportunities. The launch of ChatGPT, the most prominent AI application, is considered to be the greatest disruption in education since the start of calculators (Haque et al., 2022; Mamula Nikolić, 2024). However, this rapid integration has brought with it certain academic ethical dilemmas and questions on the real impact this technology has on students' development, engagement, satisfaction or academic performance.

The significant impact of ChatGPT in higher education is evident in how it is used, whether irresponsibly or responsibly (Hashmi and Bal, 2024) and to what extent. While AI can improve teaching and learning, it can also disrupt student motivation and focus and develop excessive resilience (Zhou et al., 2023). Various prior studies highlight the benefits of AI, such as expanding learning resources (Qadir, 2023), personalized learning (Hong, 2023), and positive impact on critical thinking and cognitive achievement (Jaboob, Hazaimah, and Al-Ansi, 2024). However, certain negative aspects and risks of ChatGPT usage have also been identified. These can be compromised academic integrity and reduced student performance (Dwivedi et al., 2023), propagation of inherent biases and hallucinations (Ivanov et al., 2024), overreliance on AI (Qawqzeh, 2024) and impeded critical thinking, evaluative and analytical skills of students (Royer, 2024; Vargas-Murillo, de la Asuncion Pari-Bedoya, and de Jesús Guevara-Soto, 2023). Particularly vulnerable are the students who have fewer interactions with their tutors (Farrelly and Baker, 2023), especially remote online students (Xu et al., 2023). To determine the impact on academic attainment most studies applied self-reported judgments (El-Seoud et al., 2023; Qawqzeh, 2024). Addressing these new challenges requires transforming not only the education system by

introducing new forms of learning and assessment (Raitskaya and Lambovska, 2024), aimed at both students and teachers.

This study aims to research the effect of ChatGPT use, and particularly the irresponsible usage of this tool, on the academic performance of university students. The study examines the relationship between ChatGPT use and academic performance, focusing on key variables such as type of study (traditional or online), instances of irresponsible ChatGPT use, and academic performance as measured by final exam points. The sample is a cohort of 116 students from the Faculty of Management at a private University in Serbia. Data is analysed using the Orange Data Mining tool, exploring different statistical models and visualising the findings.

The remainder of the paper covers a relevant Literature Review which informs the hypothesis formulation in Methodology, together with study design and analytical framework explanation. Results are featured in a separate section, followed by a Discussion exploring the main findings, theoretical and practical contributions, limitations and future research. The paper closes with Conclusions.

LITERATURE REVIEW

The consequences of ChatGPT use in education are still an open topic for academic research. The field is still nascent, posing numerous perspectives to the problem.

Benefits of ChatGPT

Many educational institutions have progressed towards using AI tools such as ChatGPT to enhance the learning environment. Customized feedback and guidance provided by these technologies significantly help students, offering quick access to knowledge (Yu, 2024; Chan and Hu, 2023; Qadir, 2023). Students often perceive ChatGPT as a “helpful friend”, providing support and guidance throughout their learning process (Šedlbauer et al., 2024).

According to El-Seoud et al. (2023), the advantages of Chat GPT as a tool in education are reflected in the quick completion of tasks and easy access to data. The authors identify various positive implications, such as increased access

to information and resources, improved student engagement, better understanding of complex topics, and personalized learning. ChatGPT and similar tools are viewed as useful research aids, helping students to generate ideas and synthesize information (Chan and Hu, 2023). Students' views on their experience with ChatGPT tend to be positive. Qawqzeh (2024) identified significant enhancements in the cognitive domain and improved learning experience. Qualitative research by Šedlbauer et al. (2024) established evidence of enhancing the critical thinking of students, although not all are enthusiastic about the use of this tool. The use of AI can be caused by the time pressure students feel, finding the efficiency of ChatGPT as a helpful solution to the issue (Abbas, Jam, and Khan, 2023; Royer, 2024).

Methods to identify the form and scale of ChatGPT's impact on students vary. Most studies measure attitudes and perceptions (Farrelly and Baker, 2023; Šedlbauer et al., 2024; Jaboob, Hazaimah, and Al-Ansi, 2024). Ivanov et al. (2024) apply the Theory of Planned Behavior to find that improving lecturers' and students' perspectives on the advantages of implementing generative AI tools could be associated with a more positive attitude, subjective norms and perceived behavioural control toward the use of such tools in education at universities. A study among university students in three Arab countries revealed that generative AI techniques and applications have a positive and significant impact on student behavior and cognitive achievement (Jaboob, Hazaimah, and Al-Ansi, 2024).

A particular aspect of research on ChatGPT benefits relates to relevance and impact on remote students. Positives of self-regulated learning techniques in terms of validated effectiveness on academic achievement in online or blended learning were identified by Xu et al. (2023). Their findings are not specific to the assistance of generative AI solutions, still, they are valuable in identifying that interactive metacognitive prompts in remote learning positively correlate with academic performance. With the aid of AI tools, remote learning can be improved with automatic tests and assignments adjusted to student's level of knowledge and learning pace (Kasneci et al., 2023).

There is a gap in existing literature in identifying, quantifying and predicting the direct impact of ChatGPT use on students' academic results. Literature also indicates varying impacts on learning depending on the level of teacher supervision and support (Qawqzeh, 2024; Kasneci et al., 2023).

Risks of ChatGPT

Despite all the advantages that are shown through numerous studies, ChatGPT also carries with it certain risks. Namely, one of the key risks faced by teachers, according to the study by Mosaiyebzadeh et al. (2023), is data privacy, which is reflected in the sensitivity of students' data when they are used for illicit purposes that do not belong to education. Then, the lack of fact-checking of content produced on demand by Chat GPT, the lack of references that Chat GPT often omits, and of course plagiarism, accidental or purposeful. Regarding the risks that concern students directly, frequently found are cheating on exams (Farrelly and Baker, 2023; Dwivedi et al., 2023; Mosaiyebzadeh et al., 2023). In cases of ChatGPT misuse, teachers face the risk of biased grading (Cotton, Cotton, and Shipway, 2024), and compromised assessment process (Eken 2023). Without crosschecking the output of ChatGPT-generated text, students risk a decrease in critical thinking and cognitive performance (Qawqzeh, 2024; Rudolph, Tan, and Tan, 2023). Students tend to express unwillingness to question the AI-generated output (Šedlbauer et al., 2024), which is a capability that should be improved. The ability to interrogate the AI output, to evaluate it and to further create from it are skills proposed to be added to Bloom's taxonomy (Ng et al., 2021).

In addition to above mentioned, specific risks are heightened in remote learning environments with online students. Lack of regular engagement with the tasks as well as with the teachers hinders students' progress and performance (Bravo-Agapito, Romero, and Pamplona, 2021). Such support from tutors is critically needed when students start using ChatGPT and other AI tools without sufficient knowledge and guidance (Farrelly and Baker, 2023).

In a study by Abbas, Jam and Khan (Abbas, Jam, and Khan, 2023) which examined the causes and consequences of ChatGPT usage among university

students, findings show that the use of ChatGPT was likely to increase tendencies for procrastination and memory loss and reduce the students' academic performance. Also, students who rely on generative AI alone may not necessarily do well in their assessments and consequently get lower grades (Lim et al., 2023). Capability for critical thinking is viewed as the key skill that higher education should nurture in students which is undermined if the use of ChatGPT is not properly guided and integrated into coursework (Royer, 2024).

ChatGPT and Ethics

The most common concerns related to the usage of ChatGPT in an academic environment are the ethics of such behaviors and academic integrity. Eken (2023) stated that the integration of ChatGPT in education can have ethical challenges, compromising the basic values of education. In certain cases, students may fail to distinguish AI-generated content from the original work or reference to other sources (Šedlbauer et al., 2024). There are still missing conventions and agreements between universities and the academic community on how to site ChatGPT content in different academic materials. Implementing practical guidelines can assist in creating informed decisions and shaping policies within educational institutions.

In two separate studies (Farhi et al., 2023; Mijwil et al., 2023) it is concluded that using ChatGPT in education has useful and concerning effects on educational integrity. Violations of scientific research and publication ethics can have serious consequences, including damage to the reputation of the researchers, journals, and institutions involved. As such, it is crucial that all researchers and publications adhere to these ethical standards to maintain the integrity of the scientific community (Mijwil et al., 2023).

Recognizing the complexities and importance of ChatGPT usage, teachers and policymakers can keep a balance by leveraging Artificial Intelligence technology to improve education while upholding ethical practices that promote critical thinking, originality, and integrity among students (Farhi et al., 2023).

126 Results from the review approach revealed that major concerns linked to ChatGPT

were in the ethics, copyright, transparency, and legal issues. Other considerations included bias, plagiarism, absence of originality, inaccurate content leading to false educational and professional narratives, limited knowledge, incorrect citations, cybersecurity susceptibilities, and the risk of spreading misinformation (Dwivedi et al., 2023; Chan and Hu, 2023; Farhi et al., 2023; Cotton, Cotton, and Shipway, 2024). There is growing concern that originality is increasingly threatened in various domains, ranging from completing homework assignments to conducting academic research (Lund et al., 2023; Rudolph, Tan, and Tan, 2023). As Chan and Hu (2023) stated, ChatGPT for brainstorming creative concepts related to education is an effective tool today. Using ChatGPT solely for generating creative ideas about education can be a positive approach as it promotes original thinking and prevents overreliance.

Recognizing that ChatGPT can be used both constructively and transparently, as well as for the purpose of deception and or substitution for individual work, Hashmi and Bal (2024) propose a useful academic typology of ChatGPT usage. The main discriminant is whether the tool is used responsibly or irresponsibly, in a transparent or opaque manner. Irresponsible AI users are classified as Savvy Cheaters in case of transparent usage, or as AI illiterate in case of opaque usage. The aim is to teach students how to use AI responsibly and transparently, understanding when and where to use AI tools and recognizing the ethical implications of AI.

There is a gap in existing literature in identifying, quantifying and predicting the direct impact of ChatGPT use on students' academic results. Literature also indicates varying impacts on learning depending on the level of teacher supervision and support (Qawqzeh, 2024; Kasneci et al., 2023). There is very little research on the direct impact of ChatGPT usage, especially under limited supervision circumstances, and online students' academic achievement.

METHODOLOGY

Considering the identified research gap, this paper aims to answer one main research question: whether the irresponsible usage of AI impacts students'

academic performance. A limited number of prior research studies explore the effects of ChatGPT use on academic performance. Even less use quantified academic achievement measured by GPA scores (Abbas, Jam, and Khan, 2023), but rely on self-reported perceptions of academic performance (El-Seoud et al., 2023; Šedlbauer et al., 2024; Qawqzeh, 2024). Applying the academic typology related to AI use proposed by Hashmi and Bal (2024b) and the proposition that irresponsible use of AI impedes academic integrity as well as learning amplification, our main research question is whether the irresponsible usage of AI impacts students' academic performance.

Hypothesis formulation

Prior research studies have mainly looked at the use of ChatGPT in an academic context from the perspective of benefits and risks for both students and teachers. Some of the identified risks are related to the propagation of inherent biases or hallucinations (Ivanov et al., 2024), or the advent of new types of plagiarism and issues with academic integrity (Hashmi and Bal, 2024; Ivanov et al. 2024). Special category of risk is related to the deskilling of students (Ivanov et al., 2024), overreliance on AI (Qawqzeh, 2024) and impeded students' independent critical thinking, analytical and evaluative skills (Qawqzeh, 2024; Royer, 2024; Vargas-Murillo, de la Asuncion Pari-Bedoya, and de Jesús Guevara-Soto, 2023). The impact of AI use on academic performance was established in previous studies, mainly measured via self-reported perceptions (El-Seoud et al., 2023; Qawqzeh, 2024) and qualitative evaluations (Šedlbauer et al., 2024). In the tests conducted, students were either asked to use ChatGPT or to openly report on their experience with ChatGPT. There is an open question on how undisclosed and inappropriate usage of ChatGPT impacts students' learning and ultimately whether it reflects on their exam scores.

Hypothesis H1: Irresponsible usage of ChatGPT negatively impacts academic performance.

128 | Another perspective on ChatGPT usage in an academic context is how it is treated by the academic institution. Some institutions encourage it, provide

relevant education for both teachers and students and support ChatGPT use intended for improving the learning experience and students' critical thinking. Supervision and appropriate training can be seen as the enablers of AI usage that will yield beneficial outcomes (Qawqzeh, 2024). Lack of it can lead to breaches of academic integrity, unethical or just ignorant use of AI tools. That can lead to lower performance on assessments and lower grades (Lim et al., 2023).

Hypothesis H2a: Online students are more likely to use ChatGPT irresponsibly.

Hypothesis H2b: Online students who use ChatGPT irresponsibly are more likely to underperform on exams.

Study design

To answer the research question, we have conducted research among students of the Faculty of Management at a private University in Serbia. The sample included all students who attended and passed the exams for three courses: Consumer Behavior, Digital Marketing and Advertising during the school year 2023/24. The total sample included 116 cases. For each case following data was recorded: type of study (traditional or online), irresponsible usage of ChatGPT for project assignment (yes or no), and points on the final exam (0–30). The rationale for these variables is provided in the following paragraphs.

The Faculty of Management offers two modes of study. The traditional mode requires students to attend lectures and to complete all their pre-exam assignments regularly during the semester. These students have frequent and direct interactions with faculty members. Online students are given access to course materials and pre-exam assignments online. They complete and submit required assignments at their own pace, without direct faculty supervision. All students have to attend the exam live. This difference in the level of interaction with faculty and ad hoc approach to covering the course material is assumed to have an impact on the quality of academic output of online vs traditional students.

One of the pre-exam assignments on each of the three selected courses is a project task. This is a written essay that is supposed to demonstrate a student's ability to solve a concrete problem, applying the knowledge taught in the course. Students select one of the topics from the list provided by the teacher. Only up to three students can apply for the same topic. The essay is about 15 pages in length. It is supposed to encompass about two-thirds of the course content. Upon essay submission, students and teachers discuss it orally, in the classroom for traditional students or online for online students. The pedagogical purpose of the project task is to prepare students for the final exam and to enable them to deploy course content in practice. The assumption is that students who take this project seriously are more likely to do better on the final exam. To summarize the reasons why this assignment was chosen to be checked for ChatGPT irresponsible usage: this is a written essay as a form; it requires students to independently, on their own, conduct research and generate content; it is broad in scope and therefore supposed to be good preparation for the final exam if done right.

Identification of irresponsible, undisclosed or opaque use of AI (Hashmi and Bal, 2024) is not a simple problem. Directly copying ChatGPT-generated text without disclosure is just one of the forms of academic integrity violations (Perkins et al., 2024). Determining whether an essay is written by a human or by LLM-based generative AI tools like ChatGPT is not easy. There are numerous tools developed for this purpose, but their reliability and accuracy are still an issue (Farrelly and Baker, 2023), especially if the content has been adapted from AI-generated output. One of the most frequently used tools for this purpose is Turnitin AI Detect (Perkins et al., 2024), despite reported issues with false positives and false negative rates (Farrelly and Baker, 2023). A recent comparison of AI detection tools (Driessen, 2024) shows that tools like Scribbr or QuillBot have the best accuracy rates and no false positives. Since QuillBot does not support the Serbian language, we have opted to use the Scribbr AI detection tool. The purpose of this check was not to determine the average use of ChatGPT in students' essays but rather to identify blunt violations of responsible ChatGPT use. Judgment on questionable content was grounded on well-known guidelines

for AI-generated text, as explained by Cotton, Cotton and Shipway (2024). We have defined this as direct copying of ChatGPT-generated text into the essay, without referencing, checking for accuracy, language check (ChatGPT does not produce fully accurate text in the Serbian language) or proper integration and interpretation of the content. Cases marked to have used ChatGPT irresponsibly were those that were found to have a significant part of the essay identified as 100% AI generated using the Scribbr AI detection.

Finally, academic performance is measured by points achieved on the final exam and not the overall grade students get. Since 70% of the grade is based on pre-exam assignments, including the project task, we wanted to record exam achievement separately from pre-exam efforts. Exams are taken in writing, requiring students to answer a set of questions for a limited amount of time. Tracking just exam points versus overall grade for the course is assessed to be a better measure of academic performance for the purpose of this research.

The conceptual model developed for this study is illustrated in Figure 1. Academic Performance is the outcome variable, while Type of Study and Irresponsible ChatGPT Usage are dependent variables. The interaction between the dependent variables assumes that the Type of Study could impact the proclivity for Irresponsible ChatGPT Usage.

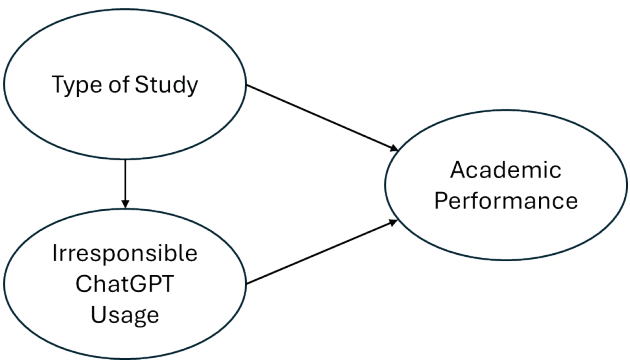


Figure 1. Study conceptual model

Analytical framework

Data were analyzed using the Orange data mining tool, version 3.37.0. „Orange is a machine learning and data mining suite for data analysis through Python scripting and visual programming“ (Demšar et al. 2013). It is one of the oldest, open-source, data mining Python-based tools. It has been widely used in various domains, including education (Abdelmagid and Qahmash 2023; Hussain et al. 2019). It is a hierarchically organized toolbox of data mining components enabling analysis like regression, decision trees, clustering, classifications, association rules, projections and others. Orange combines statistical methods and powerful data visualisation possibilities.

Results

From our data set of 116 cases, we have the following descriptive statistics about the sample, as shown in Table 1:

Table 1. Sample details

Variable	Value	Count	Percent
Type of Study	Traditional (1)	31	27%
	Online (0)	85	73%
ChatGPT Usage	Yes (1)	50	43%
	No (0)	66	57%

During the 2023/24 school year, a total of 116 students passed the Consumer Behavior, Digital Marketing and Advertising courses, 27% of them being traditional students. The data shows that 43% of students were identified to have used ChatGPT irresponsibly for their project task.

Since Exam Points was a numerical variable recording values from 0–30, its values are shown in Figure 2, using the Feature Statistics visualisation in Orange.

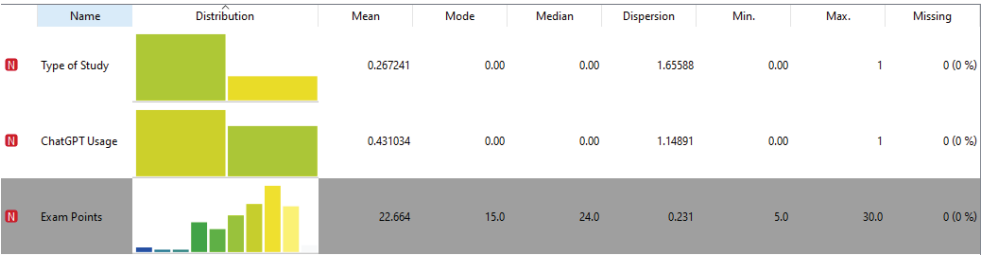


Figure 2. Feature statistics visualisation

Data in Figure 2 show that the Mean value for Exam Points is 22.664, with a minimum of 5 points and a maximum of 30 points achieved. Students needed to achieve a minimum of 15 points to pass the exam.

A comparison of mean scores on Exam Points for ChatGPT Usage is shown in Figure 3 and for Type of Study in Figure 4. Student's t-test score of 2.034 indicates a smaller difference relative to variability for ChatGPT Usage than for Type of Study (Student's t: 4.715). Calculated p values in both cases are below the 0.05 threshold meaning that there is a statistically significant difference between the two group means, although the evidence is less strong for ChatGPT Usage compared to Type of Study.

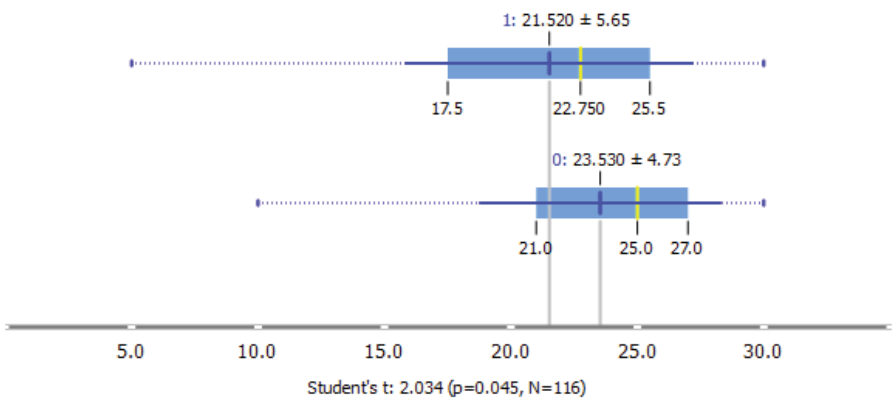


Figure 3. Exam Points means comparison by ChatGPT Usage

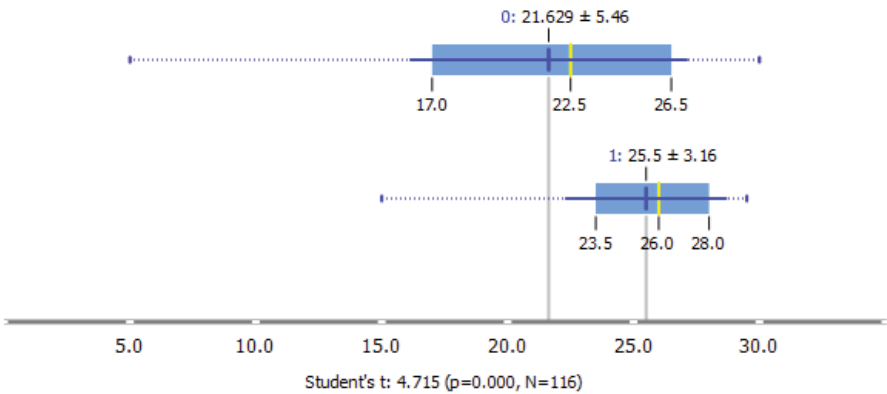


Figure 4. Exam Points mean comparison by Type of Study

These differences in data are visualised in Figures 5 and 6 via normal distribution curves. While points achieved on exams show a similar distribution curve between students who used ChatGPT and those who did not (Figure 5), there is a noticeable difference in the distribution of exam points between traditional students (red curve in Figure 6) and online students (blue curve in Figure 6).

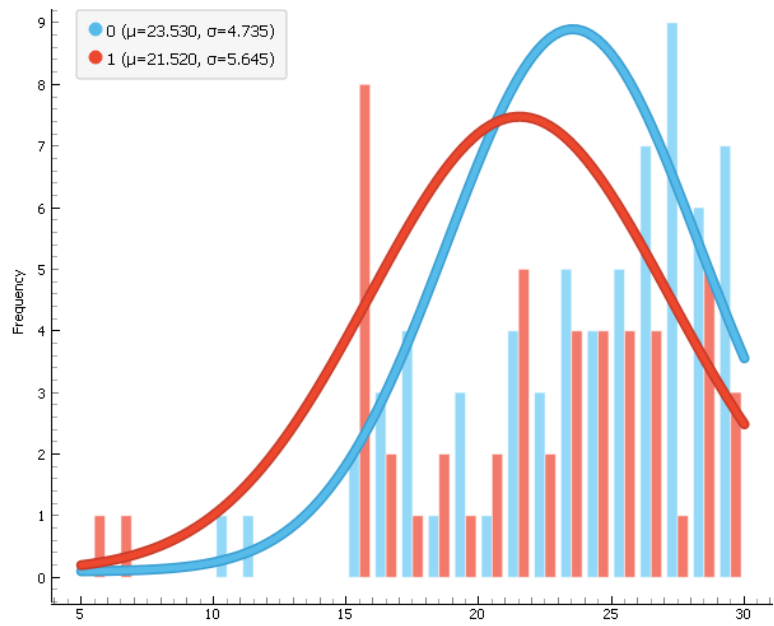


Figure 5. Exam Points Distribution by ChatGPT Usage

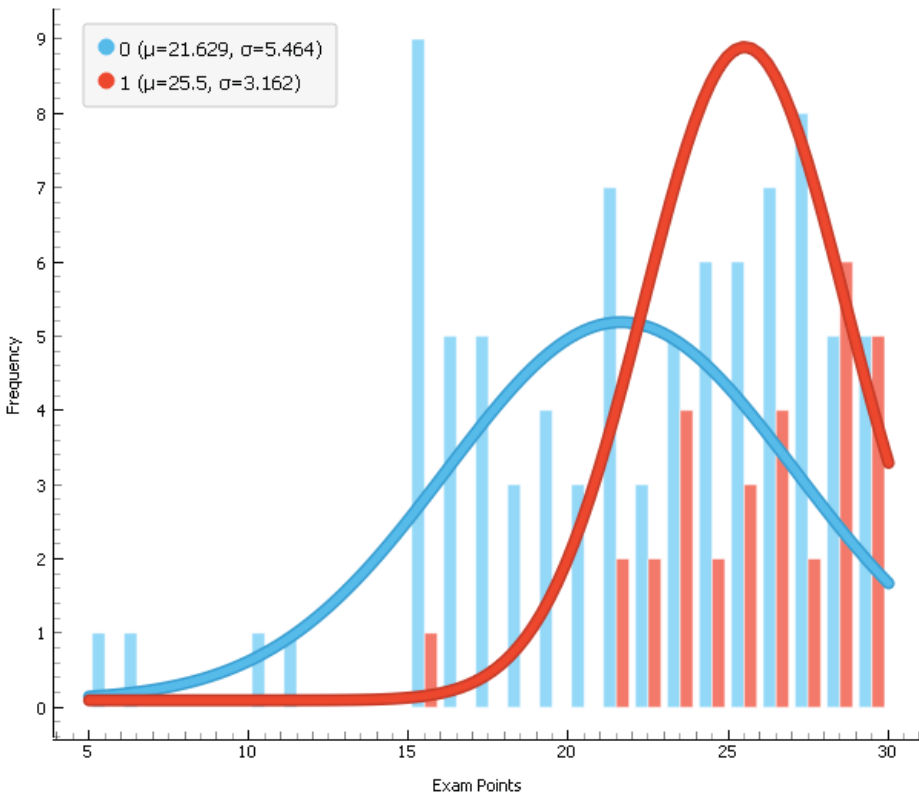


Figure 6. Exam Points Distribution by Type of Study

To measure the strength and direction of the linear relationship between variables, Pearson correlation coefficients were calculated and are shown in Figure 7. There is a moderate positive relationship between Exam Points and Type of Study. As points increase, the higher the likelihood that the student is a traditional one. The value of -0.190 suggests a weak negative correlation between ChatGPT Usage and Exam Points. The tendency for the variables to move in opposite directions is present but not very pronounced. This means that directionally if a student uses ChatGPT irresponsibly, achieved exam points will tend to be lower. With these findings, we can accept H1 that irresponsible use of ChatGPT has a negative impact on academic performance, although that impact is not very strong.

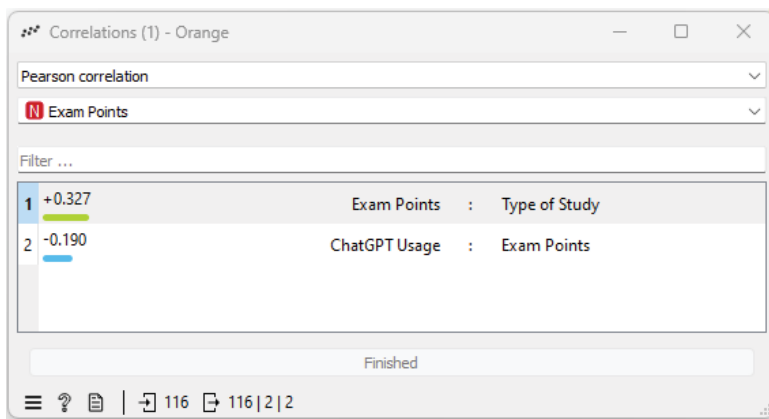


Figure 7. Pearson correlation coefficients

Another way to look at the relationships between these variables is the use of a decision tree as a classifier with a multistage approach to determining basic leadership de-complexing choices (Mohi 2020). For the Exam Points set as the target i.e. dependent variable, the decision tree hierarchy is shown in Figure 8.

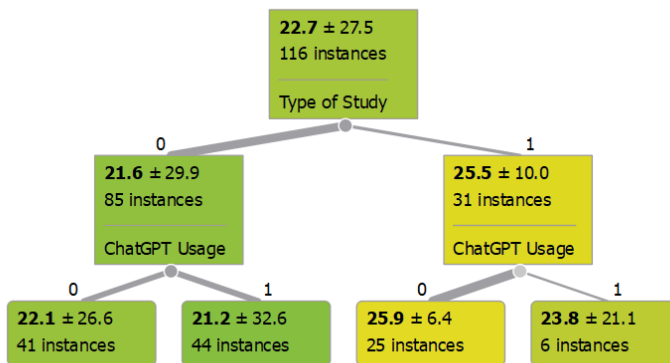


Figure 8. Decision Tree

If Exam Points are a dependent outcome variable, the Type of Study is the first sub-level of influence and ChatGPT Usage is secondary. For Type of Study 0 (online), there are 85 instances. The level below is ChatGPT Usage with almost equal split: 41 instances of 0 i.e. online students who did not use ChatGPT

irresponsibly, and 44 instances of 1 i.e. online students who used ChatGPT irresponsibly. The other side of the Decision Tree shows 31 instances of Type of Study 1 (traditional students). On the branch below ChatGPT usage shows 25 instances of 0 (not used) and only 6 instances of 1 (used). The thickness of the lines between branches indicates the significance between expected and actual instances.

Applying the k-means analysis Orange tool suggests that 4 separate clusters fully explain the data (Figure 9). These clusters are visualised in Figure 10. Darker colors indicate the tendency for lower exam scores for C1 (online students that do not use ChatGPT irresponsibly) and C2 (online students that use ChatGPT irresponsibly). The highest exam scores are in the C3 cluster of traditional students who do not use ChatGPT irresponsibly. There is finally C4: the smallest cluster of traditional students that use ChatGPT irresponsibly with diverse exam scores – there are both those with average or above average exam scores and those with below average exam scores.

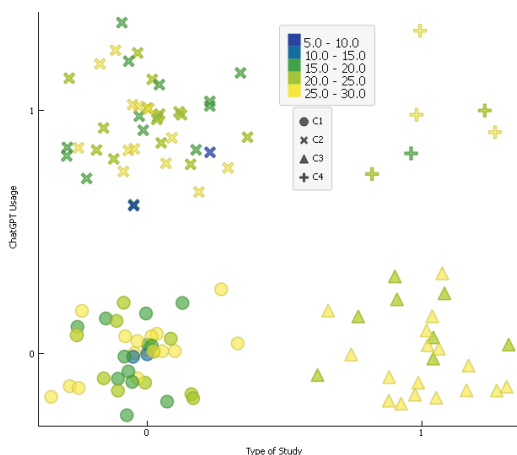
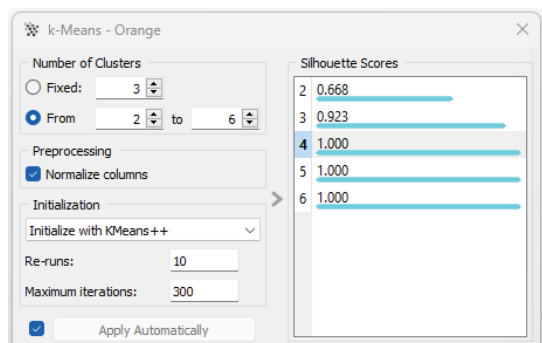


Figure 9. k-Means cluster calculation Figure 10. k-Means clusters visualisation

Identified clusters support acceptance of H2a that online students are more likely to use ChatGPT irresponsibly as well as acceptance of H2b that online students who use ChatGPT irresponsibly underperform on exams.

DISCUSSION

Main Findings

Results of the research show that 43% of students were identified to have used ChatGPT irresponsibly. Most of the prior studies have focused on the overall use of ChatGPT among university students. Research done by Fahri et al. (2023) identified that 43% of students were familiar with AI tools and that 20% were using them to complete tasks and projects. A study by Smolansky et al. (2023) found that about 29% of students in Australia and 24% of students in the US were using ChatGPT regularly for coursework. Compared to these results, not only was the ChatGPT usage higher, but it is the irresponsible usage of this tool that is this high. One of the reasons may be the significant number of online students who do not have regular supervision and interaction with the teachers. Another contributing factor may be the time period since the aforementioned papers are from 2023.

This research confirmed the impact of irresponsible ChatGPT usage on academic performance. Irresponsible usage was found to have a negative, though not strong, impact on academic performance. This finding supports previous findings by Abbas, Jam and Khan (2023) that students who often used ChatGPT for their academic assignments had poor GPAs. In the studies where academic achievement was measured by self-reported perceived progress or improvement, AI tool usage is viewed as positively impacting critical thinking and cognitive abilities (Jaboob, Hazaimah, and Al-Ansi, 2024).

An interesting finding from this research is the degree of difference between traditional and online students in the context of ChatGPT usage and consequent performance in exams. Online students significantly use AI in a way that negatively impacts their exam results and their knowledge. This supports the indicated findings of risks associated with inappropriate AI usage or overreliance on these tools (Kasneci et al., 2023; Chan and Hu, 2023; Qawqzeh, 2024). Lack of interaction with the teachers and insufficient knowledge about the appropriate

use of AI are likely to result in irresponsible ChatGPT usage, compromised academic integrity and insufficient deep engagement with required coursework (Lim et al., 2023).

Theoretical contribution

The theoretical contribution of this paper is threefold. Firstly, this is one of the few studies that uses quantitative measures of academic performance to evaluate the influence of AI tool usage by university students (Abbas, Jam, and Khan, 2023). Exam points should be a more objective, quantified proof of students' knowledge development. Secondly, this research supports prior studies which have found that irresponsible, unethical, or inappropriate use of ChatGPT negatively impacts the academic performance of university students (Abbas, Jam, and Khan, 2023; Qawqzeh, 2024). While ChatGPT can have benefits and support students' cognitive skills (Chan and Hu, 2023; Farhi et al., 2023), it has to be adequately integrated into teaching practices and supervised by the teachers. This leads to the third relevant theoretical contribution – there is a significant proclivity for online students to use ChatGPT irresponsibly and to achieve lower exam results consequently. Being online students is the primary risk for their academic performance, as the decision tree illustrated in Figure 7 shows, but they also tend to use ChatGPT irresponsibly more than traditional students. We have not found dedicated studies which explore the relationship between academic performance and ChatGPT use for online or remote students. There are suggested benefits in terms of tailored content for self-paced study when tutoring remote students (Cotton, Cotton, and Shipway, 2024; Kasneci et al., 2023) or warnings about the need for supervised remote learning (Qawqzeh, 2024).

Practical contribution

With the constant improvement of generative AI tools and their almost unlimited accessibility to university students, academic institutions need to accelerate AI adoption and implementation practices and policy-making.

Students are increasingly using generative AI tools like ChatGPT to improve their coursework (Hashmi and Bal, 2024; Šedlbauer et al., 2024) or to help them reduce the pressure and workload of assignments (Abbasi, Jam, and Khan, 2023). They should be supported to learn how to use these powerful tools responsibly, and with academic integrity. The irresponsible use of ChatGPT, which is significant as found in this study, may not be intentional, but rather arising from ignorance. The academic curriculum should provide knowledge about the use of ChatGPT and other generative AI tools.

Teachers should also continue to enhance their own ChatGPT skills. On one side, to be able to recognize its use and assess students' work accordingly (Farrelly and Baker, 2023), on the other to integrate ChatGPT into coursework in ways that can stimulate students' creativity, critical thinking, analytical and evaluative skills (Qawqzeh, 2024; Smolansky et al., 2023). Special attention should be placed on tutoring online students as they are at higher risk of academic underperformance, enforced by irresponsible ChatGPT usage. With the rising number of online students, teachers' capabilities for adequate generative AI technologies implementation in curriculum as well as appropriate engagement with students are becoming critical for academic achievement.

The data we have analyzed in this study should be relatively easy to obtain by any academic institution. If captured regularly and across the entire student population, this kind of analysis can help track students' academic achievement, progress as well as academic integrity. Based on that, universities can improve relevant aspects of tutoring, coursework and assessments.

Limitations and future research

While this study offers relevant theoretical and practical contributions, it has certain limitations. The sample size is considerable, yet limited to students that attended only three marketing courses, in one country, in one school year, at only one University in Serbia. An important limitation is the evaluation of ChatGPT usage and its classification as irresponsible. Although the Scribbr tool

was used, there was a degree of subjective assessment by the teachers on what would constitute irresponsible use in available project tasks. Application of different AI detection tools and comparison of results would improve accuracy. In addition to objective measures, students' attitudes and opinions about ChatGPT usage could be explored.

CONCLUSION

Generative AI, and especially ChatGPT, are increasingly present in today's education system. Academic research aims to understand different aspects of the usage of such tools. Different studies have identified significant positive implications on students' engagement, cognitive performance, critical thinking and academic achievements. Other studies find inherent risks of ChatGPT adoption by students, with a negative impact on critical thinking, evaluative and analytical skills, academic integrity and academic achievement. This study aimed to uncover the implications of irresponsible ChatGPT use among university students in Serbia. Findings indicate that there is a negative correlation between irresponsible use of ChatGPT and academic performance. Online students are found to be more likely to engage in such behavior and achieve lower scores on exams. These findings bring a novel understanding of the impact of AI technology on academic performance and suggest practical implications for teachers, especially with regard to support for online students.

REFERENCES

1. Abbas, Muhammad, Farooq Ahmed Jam, and Tariq Iqbal Khan. 2023. "Is It Harmful or Helpful? Examining the Causes and Consequences of Generative AI Usage among University Students." *International Journal of Educational Technology in Higher Education* 21 (1). <https://doi.org/10.1186/s41239-024-00444-7>.
2. Abdelmagid, A. S., and A. I.M. Qahmash. 2023. "Utilizing the Educational Data Mining Techniques 'Orange Technology' for Detecting Patterns and Predicting Academic Performance of University Students."

- Information Sciences Letters* 12 (3): 1415–31. <https://doi.org/10.18576/isl/120330>.
3. Bravo-Agapito, Javier, Sonia J. Romero, and Sonia Pamplona. 2021. “Early Prediction of Undergraduate Student’s Academic Performance in Completely Online Learning: A Five-Year Study.” *Computers in Human Behavior* 115: 106595. <https://doi.org/10.1016/j.chb.2020.106595>.
 4. Chan, Cecilia Ka Yuk, and Wenjie Hu. 2023. “Students’ Voices on Generative AI: Perceptions, Benefits, and Challenges in Higher Education.” *International Journal of Educational Technology in Higher Education* 20 (1). <https://doi.org/10.1186/s41239-023-00411-8>.
 5. Cotton, Debby R.E., Peter A. Cotton, and J. Reuben Shipway. 2024. “Chatting and Cheating: Ensuring Academic Integrity in the Era of ChatGPT.” *Innovations in Education and Teaching International* 61 (2): 228–39. <https://doi.org/10.1080/14703297.2023.2190148>.
 6. Demšar, Janez, Tomaž Curk, Aleš Erjavec, Črt Gorup, Tomaž Hočevar, Mitar Milutinovič, Martin Možina, et al. 2013. “Orange: Data Mining Toolbox in Python.” *Journal of Machine Learning Research* 14: 2349–53.
 7. Driessen, Koen. 2024. “Best AI Detector | Free & Premium Tools Compared.” Scribbr. 2024. <https://www.scribbr.com/ai-tools/best-ai-detector/>.
 8. Dwivedi, Yogesh K., Nir Kshetri, Laurie Hughes, Emma Louise Slade, Anand Jeyaraj, Arpan Kumar Kar, Abdullah M. Baabdullah, et al. 2023. ““So What If ChatGPT Wrote It?” Multidisciplinary Perspectives on Opportunities, Challenges and Implications of Generative Conversational AI for Research, Practice and Policy.” *International Journal of Information Management* 71 (March). <https://doi.org/10.1016/j.ijinfomgt.2023.102642>.
 9. Eken, Süleyman. 2023. “Ethic Wars : Student and Educator Attitudes in the Context of ChatGPT.” *EAI Endorsed Transactions on E-Learning* 9: 1–7. <https://doi.org/10.4108/eetel.4917>.

10. El-Seoud, Samir, Shehab Eldeen Ayman, Khaled Ahmed Nagaty, and Omar Karam. 2023. "The Impact of ChatGPT on Student Learning/Performing." In *2023 International Conference on Computer and Applications (ICCA)*, 1–5. IEEE. <https://doi.org/10.1109/ICCA59364.2023.10401713>.
11. Farhi, Faycal, Riadh Jeljeli, Ibtehal Aburezeq, Fawzi Fayez Dweikat, Samer Ali Al-shami, and Radouane Slamene. 2023. "Analyzing the Students' Views, Concerns, and Perceived Ethics about Chat GPT Usage." *Computers and Education: Artificial Intelligence* 5 (October): 100180. <https://doi.org/10.1016/j.caeai.2023.100180>.
12. Farrelly, Tom, and Nick Baker. 2023. "Generative Artificial Intelligence: Implications and Considerations for Higher Education Practice." *Education Sciences* 13 (11). <https://doi.org/10.3390/educsci13111109>.
13. Haque, Mubin Ul, Isuru Dharmadasa, Zarrin Tasnim Sworna, Roshan Namal Rajapakse, and Hussain Ahmad. 2022. "I Think This Is the Most Disruptive Technology " Exploring Sentiments of ChatGPT Early Adopters Using Twitter Data." *ArXiv:2212.05856*, 1–12. <https://doi.org/https://doi.org/10.48550/arXiv.2212.05856>.
14. Hashmi, Nada, and Anjali S. Bal. 2024. "Generative AI in Higher Education and Beyond." *Business Horizons*. <https://doi.org/10.1016/j.bushor.2024.05.005>.
15. Hong, Wilson. 2023. "The Impact of ChatGPT on Foreign Language Teaching and Learning : Opportunities in Education and Research." *Journal of Educational Technology and Innovation* 5 (March).
16. Hussain, Sadiq, Rasha Atallah, Amirrudin Kamsin, and Jiten Hazarika. 2019. *Classification, Clustering and Association Rule Mining in Educational Datasets Using Data Mining Tools: A Case Study. Advances in Intelligent Systems and Computing*. Vol. 765. Springer International Publishing. https://doi.org/10.1007/978-3-319-91192-2_21.
17. Ivanov, Stanislav, Mohammad Soliman, Aarni Tuomi, Nasser Alhamar Alkathiri, and Alamir N. Al-Alawi. 2024. "Drivers of Generative AI

- Adoption in Higher Education through the Lens of the Theory of Planned Behaviour.” *Technology in Society* 77 (December 2023): 102521. <https://doi.org/10.1016/j.techsoc.2024.102521>.
18. Jaboob, Mohammed, Manar Hazaimah, and Abdullah M. Al-Ansi. 2024. “Integration of Generative AI Techniques and Applications in Student Behavior and Cognitive Achievement in Arab Higher Education.” *International Journal of Human-Computer Interaction* 0 (0): 1–14. <https://doi.org/10.1080/10447318.2023.2300016>.
 19. Kasneci, Enkelejda, Kathrin Sessler, Stefan Küchemann, Maria Bannert, Daryna Dementieva, Frank Fischer, Urs Gasser, et al. 2023. “ChatGPT for Good? On Opportunities and Challenges of Large Language Models for Education.” *Learning and Individual Differences* 103 (February). <https://doi.org/10.1016/j.lindif.2023.102274>.
 20. Lim, Weng Marc, Asanka Gunasekara, Jessica Leigh Pallant, Jason Ian Pallant, and Ekaterina Pechenkina. 2023. “Generative AI and the Future of Education: Ragnarök or Reformation? A Paradoxical Perspective from Management Educators.” *International Journal of Management Education* 21 (2): 100790. <https://doi.org/10.1016/j.ijme.2023.100790>.
 21. Lund, Brady D., Ting Wang, Nishith Reddy Mannuru, Bing Nie, Somipam Shimray, and Ziang Wang. 2023. “ChatGPT and a New Academic Reality: Artificial Intelligence-Written Research Papers and the Ethics of the Large Language Models in Scholarly Publishing.” *Journal of the Association for Information Science and Technology* 74 (5): 570–81. <https://doi.org/https://doi.org/10.1002/asi.24750>.
 22. Mamula Nikolić, Tatjana. 2024. “THE ROLE OF DISRUPTIVE AND NON-DISRUPTIVE INNOVATION IN THE BUSINESS ECOSYSTEM.” In *3rd International Conference. „Conference on Advances in Science and Technology” COAST 2024*. Herceg Novi, Montenegro.
 23. Mijwil, Maad M, Kamal Kant Hiran, Ruchi Doshi, Manish Dadhich, and Indu Bala. 2023. “ChatGPT and the Future of Academic Integrity in

- the Artificial Intelligence Era : A New Frontier.” *Al-Salam Journal for Engineering and Technology* 2 (2): 116–27. <https://doi.org/https://doi.org/10.55145/ajest.2023.02.02.015>.
24. Mohi, Zahraa Raji. 2020. “Orange Data Mining as a Tool to Compare Classification Algorithms.” *Dijlah Journal of Sciences and Engineering*, no. 3: 13–23.
 25. Mosaiyebzadeh, Fatemeh, Seyedamin Pouriye, Reza M. Parizi, Nasrin Dehbozorgi, Mohsen Dorodchi, and Daniel Macêdo Batista. 2023. “Exploring the Role of ChatGPT in Education : Applications and Challenges.” In *The 24th Annual Conference on Information Technology Education (SIGITE '23)*, 84–89. Marietta, GA, USA. <https://doi.org/10.1145/3585059.3611445>.
 26. Ng, Davy Tsz Kit, Jac Ka Lok Leung, Samuel Kai Wah Chu, and Maggie Shen Qiao. 2021. “Conceptualizing AI Literacy: An Exploratory Review.” *Computers and Education: Artificial Intelligence* 2: 100041. <https://doi.org/10.1016/j.caeai.2021.100041>.
 27. Perkins, Mike, Jasper Roe, Darius Postma, James McGaughran, and Don Hickerson. 2024. “Detection of GPT-4 Generated Text in Higher Education: Combining Academic Judgement and Software to Identify Generative AI Tool Misuse.” *Journal of Academic Ethics* 22 (1): 89–113. <https://doi.org/10.1007/s10805-023-09492-6>.
 28. Qadir, Junaid. 2023. “Engineering Education in the Era of ChatGPT : Promise and Pitfalls of Generative AI for Education.” In *IEEE Global Engineering Education Conference (EDUCON)*.
 29. Qawqzeh, Yousef. 2024. “Exploring the Influence of Student Interaction with ChatGPT on Critical Thinking, Problem Solving, and Creativity.” *International Journal of Information and Education Technology* 14 (4): 596–601. <https://doi.org/10.18178/ijiet.2024.14.4.2082>.
 30. Raitskaya, L. K., and M. R. Lambovska. 2024. “Prospects for ChatGPT Application in Higher Education : A Scoping Review of International Research.” *Integration of Education* 28 (1): 10–21. <https://doi.org/10.15507/1991-9468.114.028.202401.010-021>.

31. Royer, Christof. 2024. "Outsourcing Humanity? ChatGPT, Critical Thinking, and the Crisis in Higher Education." *Studies in Philosophy and Education*, no. 0123456789. <https://doi.org/10.1007/s11217-024-09946-3>.
32. Rudolph, Jürgen, Samson Tan, and Shannon Tan. 2023. "ChatGPT: Bullshit Spewer or the End of Traditional Assessments in Higher Education?" *Journal of Applied Learning and Teaching* 6 (1): 342–63. <https://doi.org/10.37074/jalt.2023.6.1.9>.
33. Šedlbauer, Josef, Jan Činčera, Martin Slavík, and Adéla Hartlová. 2024. "Students' Reflections on Their Experience with ChatGPT." *Journal of Computer Assisted Learning*, no. May 2023: 1–9. <https://doi.org/10.1111/jcal.12967>.
34. Smolansky, Adele, Andrew Cram, Corina Radulescu, Sandris Zeivots, Elaine Huber, and Rene F. Kizilcec. 2023. "Educator and Student Perspectives on the Impact of Generative AI on Assessments in Higher Education." *L@S 2023 - Proceedings of the 10th ACM Conference on Learning @ Scale*, 378–82. <https://doi.org/10.1145/3573051.3596191>.
35. Vargas-Murillo, Alfonso Renato, Ilda Nadia Monica de la Asuncion Pari-Bedoya, and Francisco de Jesús Guevara-Soto. 2023. "Challenges and Opportunities of AI-Assisted Learning: A Systematic Literature Review on the Impact of ChatGPT Usage in Higher Education." *International Journal of Learning, Teaching and Educational Research* 22 (7): 122–35. <https://doi.org/10.26803/ijlter.22.7.7>.
36. Xu, Zhihong, Yingying Zhao, Jeffrey Liew, Xuan Zhou, and Ashlynn Kogut. 2023. "Synthesizing Research Evidence on Self-Regulated Learning and Academic Achievement in Online and Blended Learning Environments: A Scoping Review." *Educational Research Review* 39 (October 2020): 100510. <https://doi.org/10.1016/j.edurev.2023.100510>.
37. Yu, Hao. 2024. "The Application and Challenges of ChatGPT in Educational Transformation: New Demands for Teachers' Roles." *Heliyon*. Elsevier Ltd. <https://doi.org/10.1016/j.heliyon.2024.e24289>.

38. Zhou, Jianlong, Andreas Holzinger, Life Sciences Vienna, and Fang Chen. 2023. "Ethical ChatGPT: Concerns, Challenges, and Commandments." *ArXiv*.

CHALLENGES OF HUMAN RIGHTS PROTECTION AND SECURITY IN THE AGE OF ARTIFICIAL INTELLIGENCE

Aleksandar S. ĐORĐEVIĆ¹

Institute of Comparative Law, Terazije 41, 11000 Belgrade

Stevica DEĐANSKI²

Faculty of Social Sciences, FNP, Belgrade, Serbia

Boris JEVTIĆ³

School of Computing, RAF, Union University Belgrade, Serbia

Abstract: *Human rights protection and security are fields in which various institutional, organizational, and humanitarian actors are involved, and their effectiveness requires continuous innovation in theory and practice. In this context, this study addresses the challenges presented by the era of artificial intelligence, aiming to improve the quality of work, collaboration, and experiences of all participants in the process of human rights and security protection supported by AI technologies. The research is based on the views of 330 respondents from the Republic of Serbia, who expressed their satisfaction with the use of modern technology tools, including the latest AI solutions, through an online survey. The survey focused on matters related to the protection of human and workers' rights, as well as communication with the media, courts,*

1 djordjevic@gmail.com

2 dedjanskis@gmail.com

3 boris.jevtic10@gmail.com

companies, and public institutions. The results, analyzed through a chi-square test from a demographic perspective, indicate significant statistical differences in satisfaction levels among respondents with higher education regarding the current state of human rights protection and the actions of organizations supported by new technologies. These findings emphasize the importance of digital education for participants and organizations in the human rights and security protection process, while also warning of potential risks associated with solutions based on a lack of algorithmic transparency, cybersecurity vulnerabilities, unfairness, bias, discrimination, negative impacts on the workforce, privacy threats, and unclear accountability for possible harm. This paper contributes to the literature and future analyses, with a focus on improving decision-makers' interactions in the human rights and security protection process. It provides an impetus for greater investment in artificial intelligence and the enhancement of knowledge and skills among all stakeholders to ensure better communication, transparency, and fair solutions that guarantee the highest level of human rights and security protection.

Keywords: *Human rights protection and security, artificial intelligence, algorithmic transparency, human control, effective legislation.*

1. INTRODUCTION

Artificial Intelligence (AI) encompasses systems that exhibit intelligent behaviour by analyzing the environment and taking actions, with a certain degree of autonomy, to achieve specific objectives in a given area of application (European Commission, 2018b). AI already holds invaluable value for the global economy and is applied across various sectors of society, including regulatory and legal spheres, daily life, psychology for a better understanding of processes like memory, linguistics, and biology, as well as for learning and mastering foreign languages from a new perspective (Čurčić, Grubor, and Jevtić, 2024; Jevtić et al., 2013). These technologies contribute to creativity, the improvement

of public, business, and media services, enhanced security and quality of life, and the resolution of complex human rights protection issues. However, they also pose challenges regarding human autonomy, privacy, and fundamental rights and freedoms (OECD, 2019; Srebro et al., 2024; Špiler et al., 2023).

The integration of artificial intelligence with other technologies, including robots equipped with AI systems, is becoming increasingly significant in communication among judicial authorities, public institutions, and the economic sector, thereby enhancing the effectiveness of human rights, security, and transparency protection. Human rights protection in the context of AI-related technologies has become a notable topic in the literature (Eliot, 2022). This technology offers numerous economic and social benefits, including the capability for extensive data collection, information security, and decision-making through algorithms. However, there is concern that AI could exacerbate economic inequalities, manipulate democratic processes, and have harmful consequences for individuals and communities, threatening human rights and legitimate expectations (Mantelero and Esposito, 2021).

Technical solutions, guidelines from professional associations, standards, and regulatory mechanisms largely respond to these challenges, while media attention and public interest also play crucial roles (European Commission, 2021). The challenge posed by artificial intelligence is its scope and complexity (Russell and Norvig, 2016), as its impact is broad and present in many aspects (Makridakis, 2017).

Given these challenges, complex and multidimensional decisions based on a detailed understanding of the issues are necessary. Although various approaches are present, such as digital education, investigative journalism, and the exchange of best practices, specificity in content related to artificial intelligence is essential. This particularly pertains to numerous difficulties in the more effective implementation and practice of international human rights standards concerning AI, which can be summarized as follows:

- Lack of a comprehensive, human rights-based approach: Sector-specific approaches are more commonly adopted when applying human rights standards, focusing on certain aspects like privacy rights, while consistent application of existing safeguards across all sectors utilizing AI is rarer. Legal frameworks, where they exist, are often not implemented effectively or promptly. Additionally, reliance on large AI platforms complicates regulatory enforcement, and oversight remains fragmented.
- Insufficient transparency and information sharing: Clear and up-to-date information on AI and its potential impact on human rights is still scarce. Intellectual property protections create barriers to enforcing the right to information – including within the judiciary, national human rights frameworks, and regulatory authorities – thereby hindering independent oversight.
- Lack of proactive initiatives by member states to utilize AI to strengthen human rights: Since the majority of AI development is driven by the private sector, public authorities are less likely to adopt a proactive approach.
- Lack of algorithmic transparency, cybersecurity vulnerabilities, and threats to privacy and data protection: These issues further complicate the effective implementation of international human rights standards in the context of AI.

In this context, the role of national human rights frameworks in ensuring human rights protection through the design, development, and implementation of AI systems, as well as in strengthening independent institutional oversight, is extremely important. Promoting transparency regarding AI systems and raising public awareness about their impact on human rights, as well as adopting a proactive approach to researching AI's potential to enhance rather than undermine human rights protection, are key goals of the research presented in

this paper. Human rights and privacy protection are regulated by international and national laws, yet there is a continuous need for improving these regulations, particularly in light of new technologies like artificial intelligence. Research into literature and practice that would allow for a better understanding of these technologies' benefits, while identifying problems and vulnerabilities related to these issues, is necessary. This paper, in this regard, aims to contribute information on the perspectives of individuals as citizens, workers, and rights holders on the importance of implementing AI in decision-making processes, communicating with decision-makers, and drafting regulations and solutions for human rights and security protection, from a demographic perspective.

The paper is structured as follows: after the introduction and theoretical framework of the research, it presents the results of the empirical study, conclusions, and references used in the elaboration of the stated views.

2. THEORETICAL FRAMEWORK

United Nations, OECD, Council of Europe, European Parliament, and European Commission documents, as well as institutions such as CEPEJ and the European Data Protection Supervisor, address issues of human rights, privacy, and their protection in the context of artificial intelligence. Academic communities and civil society organizations like *Access Now* and *Privacy International* tackle the legal challenges that AI brings, including legal personhood, intellectual property, algorithmic bias, discrimination, unfairness, labour protection, privacy, data protection, and cybersecurity. The international scope of issues related to human rights, their protection, and privacy is evident in the resolutions of the United Nations (2019), OECD (2019), Council of Europe (2017, 2018, 2019), European Parliament (2017, 2018a, 2018b, 2019, 2020a, 2020b, 2020c), European Commission (2018a, 2018b, 2020), the European Commission for the Efficiency of Justice (CEPEJ) (2018), and the European Data Protection Supervisor (2016). These resolutions cover issues such as legal personhood, intellectual property

(Schenberger, 2018), algorithmic bias, discrimination, unfairness, labour protection, privacy, data protection (Wachter and Mittelstadt, 2019), security, access to justice (Raymond, 2013), algorithmic transparency, liability for harm, accountability, and oversight. Media coverage of legal issues related to artificial intelligence technologies is increasingly prevalent and specifically addresses aspects like responsibility (Mitchell, 2019), fairness in decision-making (Niiler, 2019), bias (Marr, 2019), privacy, accountability (Coldewey, 2018), privacy, and data protection issues (Meier, 2018; Williams, 2019; Lohr, 2019). Media outlets are reporting more frequently on these topics, covering aspects like accountability, fairness in decision-making, bias, and privacy protection.

Human rights and security protection in the application of artificial intelligence (AI) face significant challenges, including a lack of algorithmic transparency. This often means that it is unclear how algorithms make decisions that affect people's rights, such as denying employment, credit, or benefits, or placing individuals on no-fly lists (Cath, 2018; Desai and Kroll, 2017). The limited availability of information on algorithm functionality (Mittelstadt et al., 2016) further complicates the situation. Analyses of social, technical, and regulatory challenges highlight the need to raise awareness and strengthen oversight and accountability in algorithmic decision-making, especially in the public sector. Solutions such as algorithmic impact assessments (Reisman et al., 2018) are promising but remain in the early stages of development, and their effectiveness has yet to be fully evaluated. Transparency in communication among judicial authorities, social organizations, private companies, and the media regarding human rights and security protection also has limitations (Anani and Kraford, 2018). Cybersecurity vulnerabilities are another critical issue in the application of AI. Velser and colleagues (2017) emphasize that automated decision-making and the use of AI without human intervention increase the risk of cybersecurity attacks. The use of AI for surveillance and cybersecurity opens up new avenues for attacks, while advanced methods of targeting political messages through social media continue to evolve. Additionally, predictive policing algorithms may negatively impact citizens' fundamental

rights (Couchman, 2019). These issues pose a threat to critical infrastructure, which can have serious consequences for society and individuals, including endangering lives and access to resources. Various strategies are employed to address these issues, such as strengthening protection, implementing recovery mechanisms, involving human analysts in critical decision-making processes, risk management programs, and software upgrades (Fralick, 2019).

Deficiencies and challenges in the field of cybersecurity require the proactive and responsible use of policies, mechanisms, and tools by developers and users alike, at all stages – from design to implementation and use. When designing systems that utilize machine learning models, engineers must carefully choose architectures, consider potential attacks, and make thoughtful compromises between model complexity, explainability, and robustness (Patel et al., 2019). This balance between security and functionality is crucial for effectively addressing these challenges. Unfairness (Smit, 2017), bias (Courtland, 2018), and discrimination (Smit, 2017) often arise as problems associated with the use of algorithms and automated decision-making systems in areas like employment, credit, criminal justice (Berk, 2019), and insurance. The European Parliament report (2017) warns that algorithms used to analyze big data can lead to violations of individuals' fundamental rights and indirect discrimination against groups with similar characteristics. This can affect fairness and equal access to education and employment, as well as the evaluation of social media users. The European Parliament has called on the European Commission, member states, and data protection authorities to identify and implement measures to reduce algorithmic discrimination and bias, as well as to develop an ethical framework for transparent data processing and automated decision-making.

Various strategies have been proposed to address these issues, including regular assessments of data representativeness and algorithm adjustments to reduce bias (Danks and London, 2017). Additionally, involving humans in the decision-making process can help (Berendt, 2019), as can opening up algorithms to enable scrutiny for bias. Standards like IEEE P7003, which are being developed as part of a global initiative for ethical autonomous systems,

provide a framework for the development of algorithmic systems that avoid unjust bias. Open-source tools, such as the *AI Fairness 360 Open Source Toolkit*, enable analysis, reporting, and reduction of bias in machine learning models using advanced algorithms and metrics. To make algorithmic auditing effective, holistic, interdisciplinary approaches based on scientific and ethical foundations are necessary (Raji and Buolamwini, 2019). There are calls for greater regulatory, political, and ethical attention to fairness, particularly in protecting vulnerable and marginalized groups.

3. METHODOLOGY

3.1 Description of the study

The empirical study conducted for this paper is based on the views of 330 respondents from the Republic of Serbia during the first half of 2024. These respondents participated in an online survey, expressing their satisfaction with the use of modern technology tools, including the latest artificial intelligence solutions, concerning the protection of human and workers' rights, as well as communication with the media, courts, companies, and public institutions.

The results, analyzed through the chi-square test from a demographic perspective, indicate significant statistical differences in satisfaction levels among respondents with higher and tertiary education concerning the current state of human rights protection and the actions of organizations engaged in the human rights and security protection process in Serbia supported by new technologies. These findings highlight the importance of digital education for participants and organizations in the rights and security protection process, while also cautioning about potential risks. These risks pertain to solutions based on a lack of algorithmic transparency, cybersecurity vulnerabilities, unfairness, bias, discrimination, negative impacts on the workforce, privacy threats, and unclear accountability for possible harm. This paper contributes to the literature and future analyses, emphasizing the improvement of decision-makers' interactions in the human rights and security protection process. It encourages greater investment in

artificial intelligence and the enhancement of knowledge and skills among all actors to ensure better communication, transparency, and the delivery of fair solutions that guarantee the highest level of rights and security protection for people.

Respondents expressed their views using a three-level satisfaction scale – dissatisfied, partially satisfied, and satisfied – on statements about the availability and timeliness of information regarding regulatory measures and legislation in this area, the public system for data use permits, reporting and monitoring mechanisms for enforcement solutions, algorithm use, analytical systems for fairness assessments, complaint mechanisms, investigative journalism, as well as stakeholder dialogue and compliance with standards in the area of human rights protection.

Table 1 shows the frequencies and percentage representation of the respondents' profile elements.

Table 1: Descriptive statistics

Gender of respondents	N	%
Male	180	54.55
Female	150	45.45
Total	330	100.00
Age group	N	%
(18–29)	46	13.94
(30–45)	127	38.48
(46–60)	73	22.12
(>60)	84	25.45
Total	330	100.00
Education level	N	%
Secondary education	137	41.52
Higher or tertiary education	193	58.48
Total	330	100,00

In the empirical study, 54.55% of the total participants were male. The largest age groups were 30–45 years old and those over 60. Respondents with higher and tertiary education represented the majority, accounting for 58.48% of the sample structure.

In this research, the chi-square test method was used. Three main and three auxiliary alternative hypotheses were defined to analyze the demographic aspects of respondents' attitudes.

The hypotheses are as follows:

- H_{01} : There are no statistically significant differences between the categories of respondents' gender in terms of satisfaction with AI technologies used in human rights and security protection.
- H_{a1} : There are statistically significant differences between the categories of respondents' gender in terms of satisfaction with AI technologies used in human rights and security protection.
- H_{02} : There are no statistically significant differences between age groups of respondents in terms of satisfaction with AI technologies used in human rights and security protection.
- H_{a2} : There are statistically significant differences between age groups of respondents in terms of satisfaction with AI technologies used in human rights and security protection.
- H_{03} : There are no statistically significant differences between the education levels of respondents in terms of satisfaction with AI technologies used in human rights and security protection.
- H_{a3} : There are statistically significant differences between the education levels of respondents in terms of satisfaction with AI technologies used in human rights and security protection.

3.2 Results

Table 2 shows the contingency table with frequencies and percentages for the gender of respondents and their satisfaction with AI technologies used in human rights and security protection.

Table 2: Contingency table for gender and satisfaction with AI technologies in human rights and security protection

Frequency Overall % Column % Row %	Dissatisfied	Partially satisfied	Satisfied	Total
Male	40	78	62	180
	12.12	23.64	18.79	
	47.62	54.93	59.62	
	22.22	43.33	34.44	
Female	44	64	42	150
	13.33	19.39	12.73	
	52.38	45.07	40.38	
	29.33	42.67	28.00	
Total	84	142	104	330
	25.45	43.03	31.52	

Table 3 provides contingency frequencies, overall percentages, and breakdowns by column and row for respondents’ age groups in relation to their satisfaction with AI technologies used in human rights and security protection.

Table 3: Contingencies for age groups of respondents in relation to satisfaction with AI technologies in human rights and security protection

Frequency Overall % Column % Row %	Dissatisfied	Partially satisfied	Satisfied	Total
(18–45)	40	78	55	173
	12.12	23.64	16.67	
	47.62	54.93	52.88	
	23.12	45.09	31.79	
(>45)	44	64	49	157
	13.33	19.39	14.85	
	52.38	45.07	47.12	
	28.03	40.76	31.21	
Total	84	142	104	330
	25.45	43.03	31.52	

Table 4 provides contingency frequencies, overall percentages, and breakdowns by column and row for respondents' education levels in relation to their satisfaction with AI technologies used in human rights and security protection.

Table 4: Contingencies for education levels of respondents in relation to satisfaction with AI technologies in human rights and security protection

Frequency Overall % Column % Row %	Dissatisfied	Partially satisfied	Satisfied	Total
Secondary education	40 12.12 47.62 29.20	65 19.70 45.77 47.45	32 9.70 30.77 23.36	137
Higher or tertiary education	44 13.33 52.38 22.80	77 23.33 54.23 39.90	72 21.82 69.23 37.31	193
Total	84 25.45	142 43.03	104 31.52	330

Table 5 provides the observed and theoretical chi-square test values and statistical significances according to gender, age, and education level of respondents in relation to satisfaction with AI technologies in human rights and security protection.

Table 5: Observed and theoretical chi-square test values with statistical significance

Gender of respondents / Satisfaction with AI technologies in human rights and security protection						
Test	Chi-Square (observed)	Statistical significance (observed)	df	Chi-Square (theoretical)	Statistical significance (theoretical)	H ₀₁
Value	2.712	0.2577	2	5.991	0.05	Accept
Age of respondents / Satisfaction with AI technologies in human rights and security protection						
Test	Chi-Square (observed)	Statistical significance (observed)	df	Chi-Square (theoretical)	Statistical significance (theoretical)	H ₀₂
Value	1.144	0.5644	2	5.991	0.05	Accept
Education level of respondents / Satisfaction with AI technologies in human rights and security protection						
Test	Chi-Square (observed)	Statistical significance (observed)	df	Chi-Square (theoretical))	Statistical significance (theoretical)	H ₀₄
Value	7.296	0.0260	2	5.991	0.05	Reject

Based on the data from Table 5, the following conclusions can be drawn:

- H_{01} : There are no statistically significant differences among gender categories in satisfaction with AI technologies used in human rights and security protection.
- H_{a1} : There are statistically significant differences among gender categories in satisfaction with AI technologies used in human rights and security protection.

The significance threshold is 0.05, with a degree of freedom of 2 and a critical χ^2 value of 5.991. The observed χ^2 value of 2.712 is less than the table value of 5.991, and the observed p-value of 0.2577 is greater than 0.05. Thus, the null hypothesis is accepted, meaning no statistically significant difference was observed among the gender categories of respondents regarding satisfaction with the use of AI technologies in human rights and security protection.

- H_{02} : No statistically significant difference was observed among the age categories of respondents regarding satisfaction with the use of AI technologies in human rights and security protection.
- H_{a2} : A statistically significant difference exists between the age categories of respondents regarding satisfaction with the use of AI technologies in human rights and security protection.

The significance level is 0.05, with a degree of freedom of 2 and a critical value for $\chi^2 = 5.991$. The obtained χ^2 value of 1.144 is less than the table value of $\chi^2 = 5.991$, and the obtained p-value of 0.5644 is greater than the threshold of 0.05. Therefore, the null hypothesis is accepted, indicating that no statistically significant difference was observed among the age categories of respondents regarding satisfaction with the use of AI technologies in human rights and security protection.

- H_{03} : No statistically significant difference exists between the education level categories of respondents regarding satisfaction with the use of AI technologies in human rights and security protection.
- H_{a3} : A statistically significant difference exists between the education level categories of respondents regarding satisfaction with the use of AI technologies in human rights and security protection.

The significance level is 0.05, with a degree of freedom of 2 and a critical value for $\chi^2 = 5.991$. The obtained χ^2 value of 7.296 is greater than the table value of $\chi^2 = 5.991$, and the obtained p-value of 0.0260 is less than the threshold of 0.05. Therefore, the null hypothesis is rejected, and the alternative hypothesis H_{a3} is accepted, indicating that a statistically significant difference was observed among the education level categories of respondents regarding satisfaction with the use of AI technologies in human rights and security protection (Jevtić et al., 2024).

4. CONCLUSION

This paper provides an overview of the literature on various legal issues and challenges related to artificial intelligence (AI), including its impact on human rights (Miškić et al., 2023). The results of the study on citizens' views regarding regulatory, technological, and standardization issues of human rights protection in Serbia identify the need to improve digital literacy, particularly concerning AI, to increase the level of awareness and cooperation among all participants in the human rights and security protection process. Key actions for the protection of vulnerable social groups, which carry far-reaching social and legal implications, relate to fundamental rights such as data protection, equality, freedom, autonomy, dignity, justice, non-discrimination, and privacy. It is essential to encourage these actions through increased investments in technological infrastructure, regulatory frameworks, solutions, and the judiciary, as well as through the collaboration of all

stakeholders contributing to improving the level of protection and security in Serbia.

Notably, findings from the socio-economic research conducted as part of the SIENNA project (Jansen, 2018) further emphasize concerns regarding these issues. In addition to specific challenges, general legal issues include the lack of AI regulation in Serbia, unclear application of existing laws, and a lack of legal discourse and knowledge among judicial actors (Rodrigues, 2019). Given that AI technologies work with large volumes of data, they have multiplicative effects that further complicate legal issues and challenges in the area of human rights. The consulted literature for this paper reveals several shortcomings: at the level of policy and legislation, technical development, and in involving various stakeholders. Although some of these deficiencies have begun to be addressed, additional commitment is needed in the country. Technical deficiencies are especially significant because legal and ethical issues can be integrated into the AI system's design and development phase, which would significantly contribute to the success of AI systems.

The lack of involvement of various stakeholders poses a challenge, as their different interests (innovation, ethical compliance, profit) need to be carefully balanced. Additionally, the voices of vulnerable communities are often underrepresented. There is a need to involve a wide range of actors to consider specific vulnerabilities in different AI application contexts. As AI technologies advance, additional legal and ethical challenges will emerge, including new types of vulnerabilities that will require continuous monitoring and research. Progress in AI, robotics, and IoT will bring new dilemmas for laws in Serbia and societal values, which will require refreshed discussions and regulatory adjustments.

LITERATURE

1. Ananny, Mike, and Kate Crawford. 2018. "Seeing without Knowing: Limitations of the Transparency Ideal and Its Application to Algorithmic Accountability." *New Media & Society* 20 (3): 973–989. <https://doi.org/10.1177/1461444816676645>.

2. Berk, Richard A. 2019. "Accuracy and Fairness for Juvenile Justice Risk Assessments." *Journal of Empirical Legal Studies*. https://crim.sas.upenn.edu/sites/default/files/Berk_FairJuvy_1.2.2018.pdf.
3. Berendt, Bettina. 2019. "AI for the Common Good?! Pitfalls, Challenges, and Ethics Pen-Testing." *Paladyn, Journal of Behavioral Robotics* 10: 44–65. <https://doi.org/10.1515/pjbr-2019-0004>.
4. Council of Europe. 2019. *Unboxing Artificial Intelligence: 10 Steps to Protect Human Rights*.
5. Coldewey, Devin. 2018. "AI Desperately Needs Regulation and Public Accountability, Experts Say." *TechCrunch*, December 7. <https://techcrunch.com/2018/12/07/ai-desperately-needs-regulation-and-public-accountability-experts-say/>.
6. Cath, Corinne. 2018. "Governing Artificial Intelligence: Ethical, Legal and Technical Opportunities and Challenges." *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*. <https://doi.org/10.1098/rsta.2018.0080>.
7. Courtland, Rachel. 2018. "Bias Detectives: The Researchers Striving to Make Algorithms Fair." *Nature*. <https://www.nature.com/articles/d41586-018-05469-3>.
8. Couchman, Hannah. 2019. *Policing by Machine: Predictive Policing and the Threats to Our Rights*. Liberty Human Rights. <https://www.libertyhumanrights.org.uk/sites/default/files/LIB%2011%20Predictive%20Policing%20Report%20WEB.pdf>.
9. Čurčić, Nemanja, Aleksandar Grubor, and Bojan Jevtić. 2024. "Implementing Artificial Intelligence in Travel Services: Customer Satisfaction Gap Study at Serbian Airports." *Ekonomika* 3/2024.
10. Danks, David, and Alex J. London. 2017. "Algorithmic Bias in Autonomous Systems." In *Proceedings of the 26th International Joint Conference on Artificial Intelligence*. AAAI Press. <https://www.cmu.edu/dietrich/philosophy/docs/london/IJCAI17-AlgorithmicBias->

11. Desai, Deven R., and Joshua A. Kroll. 2017. "Trust but Verify: A Guide to Algorithms and the Law." *Harvard Journal of Law & Technology* 31: 1.
12. Eliot, Lance. 2022. "Responsible AI Relishes Preeminent Boost Via AI Ethics Proclamation by Top Professional Society the ACM." *Forbes*.
13. European Parliament. 2017. "Resolution of 14 March 2017 on Fundamental Rights Implications of Big Data: Privacy, Data Protection, Non-Discrimination, Security and Law Enforcement (2016/2225(INI))."
14. European Commission. 2018a. "Coordinated Plan on Artificial Intelligence." COM (2018) 795 final, Brussels.
15. European Commission. 2021. "Proposal for a Regulation on a European Approach for Artificial Intelligence." COM(2021) 206 final.
16. European Commission. 2020. "White Paper on Artificial Intelligence: A European Approach to Excellence and Trust." COM(2020) 65 final, Brussels.
17. Executive Office of the President. 2016. *Preparing for the Future of Artificial Intelligence*.
18. Fralick, Courtney. 2019. "Artificial Intelligence in Cybersecurity Is Vulnerable." *SC Magazine*. <https://www.scmagazine.com/home/opinion/artificial-intelligence-in-cybersecurity-is-vulnerable/>.
19. Jansen, Pieter. 2018. *SIENNA D4.1 State of the Art Review: AI and Robotics*. SIENNA Project. https://www.sienna-project.eu/digitalAssets/787/c_787382-1_1-k_sienna-d4.1-state-of-the-art-review-final-v.04.pdf.
20. Jevtić, Bojan, Stevica Dedanski, Milorad Beslać, Ratko Grozdanić, and Aleksandar Damnjanović. 2013. "SME Technology Capacity Building for Competitiveness and Export: Evidence from Balkan Countries." *Metalurgija International* 18 (special issue 4): 162–170. București: Editura Științifică F.M.R. <https://enauka.gov.rs/handle/123456789/676927>.
21. Jevtić, Bojan, Milorad Beslać, Dragana Janjušić, and Marija Jevtić. 2024. "The Effects of Digital Natives' Expectations of Tech Hotel Services

- Quality on Customer Satisfaction.” *International Journal for Quality Research* 18 (1): 1–10. <https://doi.org/10.24874/IJQR18.01-01>.
22. Lohr, Steve. 2019. “AI and Privacy Concerns Get White House to Embrace Global Cooperation.” *The New York Times*. <https://www.nytimes.com/2019/04/03/technology/artificial-intelligence-privacy-oecd.html>.
23. Mittelstadt, Brent D., Philip Allo, Mariarosaria Taddeo, Sandra Wachter, and Luciano Floridi. 2016. “The Ethics of Algorithms: Mapping the Debate.” *Big Data & Society*. <https://doi.org/10.1177/2053951716679679>.
24. Marr, Bernard. 2019. “Artificial Intelligence Has a Problem with Bias, Here’s How to Tackle It.” *Forbes*. <https://www.forbes.com/sites/bernardmarr/2019/01/29/3-steps-to-tackle-the-problem-of-bias-in-artificial-intelligence/>.
25. Mitchell, Iain. 2019. “The Use of AI Gives Rise to Huge Potential Legal Issues.” *The Scotsman*. <https://www.scotsman.com/lifestyle/iain-mitchell-the-use-of-ai-gives-rise-to-huge-potential-legal-issues-1-4924962>.
26. Makridakis, Spyros. 2017. “The Forthcoming Artificial Intelligence (AI) Revolution: Its Impact on Society and Firms.” *Futures* 90: 46–60. <https://doi.org/10.1016/j.futures.2017.03.006>.
27. Mantelero, Alessandro, and Massimo S. Esposito. 2021. “An Evidence-Based Methodology for Human Rights Impact Assessment (HRIA) in the Development of AI Data-Intensive Systems.” *Computer Law & Security Review* 41: 105561. <https://doi.org/10.1016/j.clsr.2021.105561>.
28. Miškić, Milana, Bojana Srebro, Marko Rašković, Milan Vrbanc, and Bojan Jevtić. 2024. “Key Challenges Hindering SMEs’ Full Benefit from Digitalization: A Case Study from Serbia.” *International Journal for Quality Research* 19 (2). <https://doi.org/10.22874/IJQR1902-03>.
29. Niiler, Eric. 2019. “Can AI Be a Fair Judge in Court? Estonia Thinks So.” *Wired*. <https://www.wired.com/story/can-ai-be-fair-judge-court-estonia-thinks-so>.

30. OECD. 2019. *Artificial Intelligence in Society*. Paris: OECD Publishing. <https://doi.org/10.1787/eedfee77-en>.
31. Patel, Anya, Themis Hatzakis, Kevin Macnish, Mark Ryan, and Anatolii Kirichenko. 2019. "Cyberthreats and Countermeasures." *SHERPA Project*. <https://doi.org/10.21253/DMU.7951292.v3>.
32. Privacy International & Article 19. 2018. *Privacy and Freedom of Expression in the Age of Artificial Intelligence*. <https://www.article19.org/wp-content/uploads/2018/04/Privacy-and-Freedom-of-Expression-In-the-Age-of-Artificial-Intelligence-1.pdf>.
33. Raji, Inioluwa Deborah, and Joy Buolamwini. 2019. "Actionable Auditing: Investigating the Impact of Publicly Naming Biased Performance Results of Commercial AI Products." *AAAI/ACM Conference on AI Ethics and Society*.
34. Raymond, Adam H., and Scott J. Shackelford. 2013. "Technology, Ethics, and Access to Justice: Should an Algorithm Be Deciding Your Case?" *Michigan Journal of International Law* 35: 485.
35. Russell, Stuart J., and Peter Norvig. 2016. *Artificial Intelligence: A Modern Approach*. Pearson Education Limited.
36. Rodrigues, Rowena, Alexandros Panagiotopoulos, David Wright, Themis Hatzakis, and Stella Laulhé Shaelou. 2020. "SHERPA Deliverable 3.3 Report on Regulatory Options." *SHERPA Project*. <https://doi.org/10.21253/DMU.8181827.v2>.
37. Schönberger, Daniel. 2018. "Deep Copyright: Up-and-Downstream Questions Related to Artificial Intelligence (AI) and Machine Learning (ML)." *Zeitschrift für Geistiges Eigentum/Intellectual Property Journal* 10 (1): 35.
38. Srebro, Bojana, and Bojan Jevtić. 2024. "Improving Decision-Making Efficiency Through AI-Powered Fraud Detection and Prevention." *International Congress on Project Management of ICT*, Arandelovac, 2024.

39. Smith, Laura. 2017. "Unfairness by Algorithm: Distilling the Harms of Automated Decision-Making." *Future of Privacy Forum*. <https://fpf.org/2017/12/11/unfairness-by-algorithm-distilling-the-harms-of-automated-decision-making/>.
40. United Nations. 2019. *United Nations Activities on Artificial Intelligence (AI)*. <http://handle.itu.int/11.1002/pub/813bb49e-en>.
41. Williams, Holly. 2019. "Big Brother AI Is Watching You." *IT ProPortal*. <https://www.itproportal.com/features/big-brother-ai-is-watching-you/>.
42. Wachter, Sandra, and Brent D. Mittelstadt. 2019. "A Right to Reasonable Inferences: Rethinking Data Protection Law in the Age of Big Data and AI." *Columbia Business Law Review*. https://oralex.ac.uk/objects/uuid:d53f7b6a-981c-4f87-91bc-743067d10167/download_file?file_format=pdf.

APPLICATION OF ARTIFICIAL INTELLIGENCE IN DETECTION AND PREVENTION OF PEER VIOLENCE: A LITERATURE REVIEW

Zorana Bodiřoga¹

Faculty of Political Sciences University of Belgrade

Abstract: *Peer violence in schools is a serious social problem that can have long-lasting and detrimental effects on the mental health of children and youth. Recently, the application of artificial intelligence (AI) has been on the rise across all sectors of society, with modern research indicating its potential use in the detection and prevention of peer violence. The aim of this research is to review the literature on peer violence with a special focus on the possibilities of implementing AI in detecting and preventing this social issue. Through a systematic review of the literature, current and relevant studies addressing peer violence, as well as those exploring the use of AI in education, specifically in the detection and prevention of peer violence, were analyzed. The research findings show that AI can significantly contribute to the early detection of indicators pointing to peer violence through the analysis of text, speech, social interactions, and social media. Additionally, the main advantages and challenges of applying AI were identified, including legal and ethical issues and cultural sensitivity.*

The conclusion of the paper emphasizes the importance of an interdisciplinary approach in the implementation of AI for combating peer violence, as well as the need for further research to enable more effective use of such solutions in real-world conditions.

Keywords: *peer violence, artificial intelligence, prevention, literature review.*

INTRODUCTION

Peer violence in schools is a serious social problem that can have long-lasting and detrimental effects on the mental health of children and youth. According to research conducted in the United States, about 20% of high school students experience some form of peer violence, and about 70% of students report witnessing some form of violence in school (Nansel et al., 2001). Globally, an average of 10–30% of adolescents have reported experiencing some form of violence (Cook et al., 2010; Schoeler et al., 2018). A study on school violence from 2014 shows that about 30% of elementary school students experienced violence at least once during their schooling, with the percentage increasing in higher grades (Popadić et al., 2014). It is important to note that the reason for recording higher percentages in higher grades may not necessarily be an increase in violence, but rather that older students can better detect different forms of violence.

Peer violence can manifest in various ways, including physical violence, verbal harassment, social exclusion, and the like. Recently, digital violence has also become an increasingly relevant form of peer violence. According to the Law on the Basics of the Education System, digital violence and abuse involve the misuse of information and communication technologies that can result in harm to another person and endanger dignity (ZOSOV, 2023). Peer violence can have serious consequences for the mental health of children and youth, including stress, anxiety, depression, increased risk of substance abuse, and suicide (Bradshaw et al., 2015), making the development of effective strategies

for preventing peer violence essential. Research has shown that a strengths-based approach can be very effective in preventing peer violence. This approach focuses on identifying and strengthening existing strengths and resources in children and youth, rather than their deficiencies and weaknesses.

Modern research also demonstrates the effectiveness of AI-based resources in detecting and preventing peer violence. Techniques of deep learning and natural language processing (NLP) have been proposed for real-time violence detection from surveillance footage, including identifying firearms, robbery, fights, sexual harassment, and fires in video recordings (Davidson et al., 2017). AI-based mechanisms, such as automatic speech recognition (ASR), have been proposed for addressing domestic violence cases during the pandemic, where physical access to authorities was limited (Davidson et al., 2017). Additionally, real-time violence detection systems using deep learning on unmanned aerial vehicles (UAVs) have been developed, improving requirements such as computational speed, model capacity, accuracy, and object tracking technology (Pham Van, 2023). These AI-based approaches also offer efficient, fast, and precise solutions for detecting and preventing violence, enabling timely intervention and ensuring the safety and security of individuals and society as a whole. With the return of students to schools after the global pandemic, school associates, parents, and students have faced new challenges, motivating researchers to apply AI-based approaches to detect and prevent peer violence in schools.

PEER VIOLENCE

Recently, increasing attention has been given to preventive programs aimed at reducing the incidence of this type of violence among youth. One approach based on the strengths perspective involves implementing information technologies and artificial intelligence (AI) programs that facilitate the easier detection of various forms of peer violence, thereby leading to more effective prevention.

Peer violence is a serious social problem occurring worldwide. Although violence can be defined as the negative behavior of one person that endangers the safety, physical or mental health, and dignity of another person, broader and narrower definitions of this phenomenon can be found in the literature, indicating a very extensive denotation of the term. A broader definition of violence is given by Milosavljević (1998), who defines it as actions and behaviors of an individual or group, social institution, or organization, which involve the use of force – whether psychological, physical, or of another kind – that endangers human integrity and leaves consequences.

Olweus defines violence as persistent negative behavior by one or more individuals directed at another person with the aim of endangering safety or rights (Olweus, 1997). Similarly, with the exception of the persistence aspect, the Law on the Basics of the Education System defines violence as “any form of verbal or non-verbal behavior, whether single or repeated, that results in actual or potential endangerment of the health, development, and dignity of a child, student, or adult.” The same definition is found in the Manual for the Implementation of the Special Protocol (2009), which also presents a classification of violence according to its form:

1. Emotional/psychological violence – immediate or prolonged endangerment of the child’s psychological or emotional health;
2. Physical violence – actual or potential physical injury to the child/student;
3. Electronic violence – misuse of information technologies;
4. Sexual violence – involving children in sexual activities they do not understand and are not developmentally ready for;
5. Social violence – excluding a child or children from the group (Manual, 2009).

Peer violence occurs when violence happens between children of the same or similar age. Jerković (2010) adds that it is possible to detect an unequal power relationship in the occurrence of peer violence, meaning that the child subjected

to peer violence is weaker and more helpless compared to the peer or group of peers, while the literature on peer violence additionally emphasizes persistence. Based on this, certain characteristics and a definition of peer violence can be derived for further research purposes:

1. Intent to endanger the physical, psychological, or emotional health of another child or group of children of the same or similar age.
2. Actual or potential endangerment of the physical, psychological, or emotional health of another child or group of children of the same or similar age.
3. Peer violence involves an unequal power relationship or imbalance of power (Manual, 2009), where the victim is weaker and more helpless, and the perpetrator is more powerful and stronger.
4. Peer violence most often implies continuity, i.e., the persistent endangerment of the physical, psychological, or emotional health of another child or group of children of the same or similar age.

Based on these characteristics, we can deduce the definition of peer violence for further research purposes: Peer violence is intentional and persistent behavior by an individual or group aimed at actually or potentially endangering the physical, psychological, or emotional health of another child or group of children of the same or similar age, and can be carried out verbally, through physical force, psychological or social treatment of an individual or group, or by using information technologies.

Thus defined, peer violence is present in all social strata and cultures and has serious consequences for the health and well-being of children and youth. According to UNICEF research, peer violence occurs in almost all countries worldwide. It is most prevalent in South American countries, Central and Eastern Europe, as well as in Asia and Africa (UNICEF, 2017). In some countries, such as the USA and Canada, it is estimated that peer violence affects nearly 30% of children and youth, as confirmed by WHO research from 2018. Data on peer

violence in Serbia also aligns with global research. A study on school violence from 2014 shows that about 30% of elementary school students experienced violence at least once during their schooling (Popadić et al., 2014).

In recent years, the problem of peer violence has become increasingly prominent in public and media discourse, which does not necessarily mean that peer violence is more prevalent, but may indicate positive outcomes of initiatives and actions aimed at reporting violence, as well as positive outcomes of educating students, teachers, and professional associates about this problem to ensure adequate response and better detection of various forms of peer violence.

According to the research by Popadić and associates (2014), boys are more often involved in physical violence, while girls are more often subjected to verbal and psychological violence. The most common forms of peer violence in Serbia are insults, belittlement, spreading rumors, exclusion from groups, as well as physical violence such as hitting and pushing. An increasingly prevalent form of peer violence is electronic violence, to which girls are particularly exposed.

At the end of the 20th century, the phenomenon of peer violence began to be approached from a group perspective, recognizing that in situations of peer violence, it is not possible to isolate solely the victim or the perpetrator of violence. Initial research showed that six roles could be identified in an act of peer violence: victim, perpetrator, accomplice, instigator of the perpetrator, observer (outsider), or defender of the victim (Salmivalli et al., 1996). Subsequent research further expanded this list, so in Olweus' work, we find eight roles depending on whether they have a positive, neutral, indifferent, or negative attitude towards peer violence, or whether they are direct participants or not (Olweus, 2001). This research indicates that peer violence is not solely a problem of the perpetrator and the victim but involves a larger number of subjects with different roles and participation in this issue, opening the door to ecological systems theory in a sociological sense, i.e., an approach that encompasses a larger number of participants and their interpersonal connections.

176 The consequences of peer violence in Serbia can be very serious, including depression, anxiety, suicidal thoughts, decreased academic achievement, increased absenteeism from school, and health problems, while the negative

impact affects not only the victim of peer violence but also the perpetrator, their families, and society as a whole.

METHODOLOGY

To conduct a thorough and systematic review of the available literature on the application of artificial intelligence (AI) in the detection and prevention of peer violence, ensuring a high level of rigor and objectivity in the analysis and synthesis of collected data, diverse sources of information relevant to the topic were included according to established criteria.

The following databases and search tools were used to identify relevant literature: Web of Science, Scopus, Google Scholar, PubMed, IEEE Xplore, as well as specialized educational databases such as ERIC (Education Resources Information Center). Searches were conducted using keywords and phrases, and their combinations, such as “artificial intelligence and bullying prevention”, “machine learning in education”, “detection of cyberbullying”, “information technology and school bullying”, “AI school bullying detection”, “school bullying prevention”, The search also included Serbian translations of these terms and phrases.

For the literature review, works that directly address the topic of peer violence and the application of artificial intelligence, with a particular focus on the detection and prevention of violence, were selected. The literature included a wide range of academic works, including articles published in peer-reviewed journals, books, doctoral dissertations, and conference papers, as well as reports from relevant international and national organizations.

The search yielded 129 research and theoretical papers, after which criteria were applied to include the works in the literature review. Included in the review were those papers specifically focused on the application of artificial intelligence in the domain of peer violence detection and prevention, as well as those published from 2014 to 2024 to ensure the relevance and currency of the data. Exclusion criteria for the literature involved insufficient focus on artificial intelligence or peer violence. Additionally, works published before 2014 and

those lacking empirical evidence to support their conclusions were excluded from the review.

Only 13 papers met the established criteria from which key information, methods, findings, and conclusions were extracted. The analysis also included a critical review of the methodologies used in the studies, an assessment of the validity and reliability of the findings, and a discussion of the potential implications for practice and future research

APPLICATION OF ARTIFICIAL INTELLIGENCE IN THE DETECTION AND PREVENTION OF PEER VIOLENCE

While the topic of peer violence is globally widespread and thoroughly researched, its relevance persists. The application of AI-based tools has only seen significant expansion in the past decade, limiting the scope of research on this topic. A study by Topcu-Uzer and Tanrikulu (2018) highlighted that information technology tools are not yet capable of reliably and consistently detecting and preventing cyber or other forms of violence. Until 2020, the number of studies indicating the effectiveness of new technological solutions against peer violence and cyberbullying was very limited, with exceptions in systems like text-based detection (Soundar and Ponesakki, 2016), participant vocabulary coherence (Raisi and Huang, 2016), or morphosemantic patterns (Ptaszynski et al., 2016). Moreover, recent studies, such as the Rethink software and empathy-increasing video materials, are being conducted to combat these issues. These will be key elements in developing new technological solutions to fight bullying and cyberbullying and in future research demonstrating their effectiveness.

In 2018, Silva et al. conducted a literature review focusing on interventions aimed at preventing peer violence in schools. Although the study did not analyze the effectiveness of AI and digital tools, the conclusions can be applied to peer violence prevention in general. Silva et al. classified interventions into four categories: multi-component, involving the entire school, social skills training, academic within the curriculum, and computerized. Each category was

thoroughly researched in relation to relevant research databases, indicating a high effectiveness rate for interventions involving multiple components and the entire school community, and those adapted to the cultural and social factors of all school community members (Silva et al., 2018). This suggests the necessity for implemented solutions to be tailored to the school community environment and the characteristics of students to establish better intervention effectiveness.

However, certain studies from 2014 indicate the implementation of modern solutions in the field of information technology, such as the detection of emotional states using smartphones in controlled environments for detecting peer violence in schools (Hany et al., 2014). By classifying emotions in the form of frequencies displayed on an electrocardiogram, this research helped establish a connection between the emotions of participants and the MAHNOB database using a smartphone and sensors (Hany et al., 2014). This laid the groundwork for detecting fear, power, and other emotions significant for detecting the prevalence of peer violence among children and youth, paving the way for the implementation of AI-based solutions for these purposes.

The primary use of artificial intelligence was observed in the application of machine learning algorithms to analyze patterns in data collected from schools. For instance, Yan et al. (2023) used machine learning to identify risk factors and prevent victimization resulting from peer violence among adolescents. This study used algorithms to analyze data from over 410,000 students, contributing to high accuracy in outcome prediction. Kumar et al. (2021) developed an AI-based system using neural networks, image processing, and natural language processing to monitor bullying situations in primary and secondary schools. This system integrates data from various sources, such as surveillance cameras, social media, and school records, and it successfully identified victims and perpetrators of peer violence with an accuracy of 87%.

Cedillo et al. (2022) reviewed literature published from 2009 to 2021 to offer a comprehensive overview of tools and methods in information technology used in the detection and prevention of physical and cyberbullying. From this perspective, the time period of the included studies is crucial for observing

the issue of peer violence, especially in virtual environments, considering the possibility of comparing data before and after the COVID-19 pandemic, during which many social interactions migrated from the physical world to virtual environments, contributing to an increase in cyberbullying incidents. Various methods were already used during this period in combating peer violence within the realm of information technology, such as video surveillance systems, serious games, and the use of artificial intelligence on distance learning platforms, all aimed at educating children and youth about this global problem, as well as providing monitoring and intervention capabilities.

The research highlighted a significant increase in technological solutions against peer violence but also a lack of studies confirming the effectiveness of these tools. Regarding artificial intelligence, perhaps the biggest challenge has been detecting elements of peer violence in written and verbal communication due to the inability to determine the appropriate context, which can lead to false results. Nonetheless, even without fully effective AI applications, the research has produced benefits from using IT tools in establishing engaging and educational programs that reduce the incidence of peer and cyberbullying. Therefore, this literature review serves as an excellent starting point for further research on the application of AI and digital tools in the detection and prevention of peer violence, emphasizing the need for empirical studies to examine the effectiveness of these tools and methods.

Milošević, Van Royen, and Davis (2022) investigated the application of AI in relation to cyberbullying, harassment, and abuse in terms of managing content that could harm students. This study particularly highlighted the complexity of peer violence in the virtual space, requiring a multidisciplinary approach that encompasses not only information technology but also psychology, sociology, and legal frameworks. The study specifically points to the frequent inability to restrict content, thereby posing challenges and difficulties in this regard. Consequently, peer violence perpetrated in this manner is often more

complex in terms of classification and ambiguity, potentially leading to multiple consequences for the victim's psychosocial and emotional well-being.

This complicates the detection of peer violence, considering AI's inability to recognize peer violence in language when irony, sarcasm, or ambiguous word meanings are used, which can lead to false positives in violence detection. Moreover, content in the virtual world can be transmitted in various forms, such as images, sound, text, or video, further affecting the meaning and context of the conveyed message. Finally, the challenge in content management also lies in the volume of content available on online platforms and social networks, with the number of posts continuously increasing in real time.

The authors mention several AI-based techniques that can be used for content management aimed at detecting peer violence. Natural language processing and machine learning can be automated to detect content that violates specific regulations, such as automatically removing images depicting pornographic content or posts containing derogatory language. Deep learning models can contribute to understanding the context and usage of words in a text, enhancing the precision and accuracy of detecting content that can be classified as peer violence. The research emphasizes the importance of an interdisciplinary approach in implementing these AI-based tools to prevent ethical issues related to the misuse of protected data, thereby providing protection for violence victims. Therefore, while the significance of using such tools is unquestionable, conditions must be met to ensure their proper implementation.

In a 2023 study, Milošević et al. continued researching these issues, analyzing the role of AI in combating peer violence from the perspective of adolescents and children, aligning with Silva et al.'s (2018) appeal for implementing measures and interventions that consider the needs of the research subjects. This study (Milošević et al., 2023) focused on addressing challenges identified in the previous year's research, such as children's right to privacy, protection, freedom of speech, and other ethical issues according to the United Nations Convention on the Rights of the Child (1989). The research was conducted in Ireland with children aged 12 to 17 who participated in interviews

and responded to questions within hypothetical scenarios involving AI-based content management strategies and their impact on their rights. While many participants generally accepted AI-based content management, many also expressed concerns about their privacy, particularly regarding private message exchanges and methods involving facial recognition. Participants also worried that such interventions might judge their communication with peers, potentially misinterpreting their intentions to joke in situations that AI might recognize as peer violence.

Thus, the implementation of AI-based tools necessitates respecting the privacy and personal data of the subjects to ensure the effectiveness of these tools.

Similar ethical issues were found in the research by Kumar et al. (2020). The advantage of Kumar et al.'s (2020) research is that it used existing infrastructure in many schools, such as video surveillance. The study focused on implementing advanced machine learning techniques in school surveillance systems to effectively detect peer violence. Using convolutional neural networks to analyze video recordings from surveillance cameras and long short-term memory to examine patterns in video and audio recordings, the research involved recognizing patterns related to peer violence from databases of surveillance cameras, student forums, platforms, and school records (attendance, logs, reports).

Indicators analyzed included physical violence, verbal abuse, social exclusion, and changes in academic performance and school activity attendance. The analysis also included assessing the emotional tone of written communication on student forums and platforms, contributing to an 87% accuracy in classifying situations and forms of peer violence. With this precision and automated real-time event monitoring via video surveillance and communication tracking, the research indicated the possibility of timely intervention both in schools and the virtual spaces used by students.

182 This study also highlighted the problem of context, specifically potential errors in detecting peer violence when using natural language in jokes, sarcasm,

irony, or other forms of statements that could be misinterpreted as peer violence without adequate context. Additionally, Kumar et al. noted the necessity of improving and implementing facial recognition to identify perpetrators and victims of peer violence, which potentially raises further ethical issues concerning biometric data and data protection.

Video surveillance was also used in the study by Orru et al. (2023). This research analyzed the “BullyBuster” project, which integrates AI, video surveillance, and psychological theories to implement a multidimensional system aimed at detecting and preventing peer violence in both physical and virtual environments. For example, it enables the detection of potential peer violence incidents through video surveillance and text analysis on social media and other platforms by detecting patterns. An addition compared to previous research is the analysis of typing patterns to detect potential aggression, and a module is implemented to detect content that could potentially harm a student’s safety or reputation. The project showed promising results, especially due to the implementation of protocols and measures compliant with European laws and regulations regarding data protection, further highlighting the importance of these ethical and legal issues in AI implementation.

Machine learning was also implemented in a study conducted in China by Yan et al. (2023) to identify risk factors for the occurrence of peer violence, both in physical and virtual environments. With the aim of early detection of potential peer violence and its risk factors, the study aimed to prevent peer violence by enabling timely intervention before it occurs. Using six different machine learning algorithms, the study collected and analyzed data from over 410,000 students, comparing them across more than 40 variables. The wide range of these variables (academic performance, personal and social factors, psychological and physical factors, etc.) enabled a holistic approach to the problem of peer violence, identifying the factors and indicators that most frequently contribute to its occurrence.

The “Random Forest” and “Light Gradient Boosting Machine” algorithms demonstrated the highest accuracy in predicting situations that could be detected as

peer violence. They also highlighted particularly risky social factors contributing to peer violence among victims, such as mental health issues, physical health problems and deformities, disrupted family or social dynamics, and living in inadequate conditions. These algorithms facilitated identifying individuals and classifying them as high-risk or low-risk for potential victimization, allowing not only timely reaction and preventive measures before violence occurs but also timely protection of victims, thus safeguarding the safety and emotional well-being of vulnerable students.

However, the research also points out the challenge posed by the large volume of data, algorithms, and the wide range of variables. With such extensive data, establishing precise causal links between risk factors and events that can be considered peer violence becomes difficult, complicating early detection and prevention. This problem could be addressed with more adequate classification following long-term model testing, which could also contribute to targeted interventions and preventive measures based on accurately detected risk factors.

Recently, the use of chatbots, or virtual assistants, has become more common. Lafrance and Villeneuve (2024) investigated the effectiveness of chatbots in combating school bullying. What sets chatbots apart from other tools is their interactivity and real-time response capability, given their ability to reply immediately to questions and their accessibility since chatbots are implemented in all smart devices. The study showed general acceptance among children to communicate with chatbots, primarily because they don't fear judgment and can freely express their thoughts and feelings. However, the challenges with chatbots stem from their advantages: the lack of a human factor in detecting emotions and states, which could significantly enhance the detection of peer violence. Additionally, chatbots may provide generic responses that are not entirely suitable for the student's situation, potentially increasing emotional stress instead of reducing it. However, the study suggests that chatbots are underutilized, considering that they, with minor improvements in emotional intelligence, can also generate valuable data for further research on peer violence and establish effective prevention programs.

Emotional intelligence and empathy are precisely human advantages over artificial intelligence, highlighting the importance of the research by Kim et al. (2021). Kim et al. explored current approaches to detecting cyberbullying through the lens of human emotions, states, and beliefs. The focus was on machine learning algorithms, and over 56 studies on this topic noted that the effectiveness of tools for combating peer violence largely depends on human parameters. For example, cyberbullying detection models do not include theoretical behavioral models or knowledge of the psychosocial dynamics of peer violence but rely solely on technical aspects, such as engineering scenario modeling. Additionally, anti-bullying models rarely validate data obtained through AI, affecting the effectiveness of the tools. Finally, there is a lack of long-term empirical studies considering not only the sustainability of these models but also managing consequences such as false positives, classification issues of peer violence, or ethical problems arising from personal data monitoring.

In other words, Kim et al. systematically highlight the necessity of including the human factor in implementing artificial intelligence (AI) in the detection and prevention of peer violence. This is due to humans' ability to detect emotional states and nuances in contexts where AI is applied, without excluding the benefits AI brings to combating peer violence.

CONCLUSION

The results of this research indicate that AI has significant potential in detecting and preventing peer violence. Automation and the application of AI-based technologies, such as natural language processing and machine learning, open up the possibility of timely detection of potential factors that can lead to peer violence. Unlike human factors, the use of AI allows for the processing of large amounts of data and rapid real-time detection. Early interventions are also facilitated by analyzing social interactions, both in physical and virtual environments, by AI-based technologies, which can contribute to classifying interactions and identifying risk factors. Additionally, the research pointed out

the shortcomings of AI applications, such as still inadequate identification of emotional content and personalized interventions when it comes to peer violence, which is the first step toward improving these functionalities.

However, research on the application of AI is consistent regarding potential challenges, such as ethical and cultural issues. The uniform stance of researchers and study subjects is that the application of AI carries ethical dilemmas, especially in the realm of privacy, data protection, and adequate content detection. Cultural factors can also threaten the effectiveness of AI applications, as cultural and generational differences can affect the appropriate interpretation of interactions, leaving room for false results.

One advantage of contemporary times is the easier integration of AI with existing educational systems and methods of social interaction (distance learning platforms, video surveillance, social networks, etc.), providing a large database for creating a comprehensive picture of the social dynamics within a school community. This can also contribute to monitoring changes in student behavior over time, enabling easier identification of trends and changes, and thus the easier implementation of measures and strategies to prevent peer violence. While these results show promising applications of AI in combating peer violence, it is essential to emphasize the need for additional empirical research to refine existing tools and gather information on the sustainability of such models and their long-term effects on students' psychosocial health, in their best interest.

REFERENCES

1. A Familiar Face: Violence in the lives of children and adolescents. UNICEF. New York. 2017.
2. Bradshaw, C.P., Waasdorp, T.E., Leaf, P.J., & Smith, J.D. (2015). A cluster-randomized trial of the KiVa anti-bullying program in primary schools: Does it work? *Journal of Educational Psychology*, 107(4), 1016–1035.
3. Cedillo, P., Bermeo, A., Betancourth, A., Espinosa, F., Illescas, L. & Jadán, J. (2022). A Systematic Literature Review on Technological Solutions to Fight Bullying and Cyberbullying in Academic Environments.

- In Proceedings of the 14th International Conference on Computer Supported Education (CSEDU 2022) – Volume 1, pp. 413–420. <https://doi.org/10.5220/0011091800003182>
4. Cook, C. R., Williams, K. R., Guerra, N. G., Kim, T. E., & Sadek, S. (2010). Predictors of bullying and victimization in childhood and adolescence: A meta-analytic investigation. *School Psychology Quarterly*, 25(2), 65–83. <https://doi.org/10.1037/a0020149>
 5. Davidson, T., Warmesley, D., Macy, M., & Weber, I. (2017). Automated Hate Speech Detection and the Problem of Offensive Language. Proceedings of the International AAAI Conference on Web and Social Media, 11. <https://doi.org/10.1609/icwsm.v11i1.14955>
 6. Duc-Anh, Pham. (2023). A novel violence detection for drone surveillance system. 1–6. <https://doi.org/10.1109/IC3S57698.2023.10169405>
 7. Ferdinando, H., Ye, L., Seppänen, T., & Alasaarela, E. (2014). Emotion Recognition by Heart Rate Variability. *Australian Journal of Basic and Applied Science*, 8(14), 50–55. ISSN 1991–8178
 8. Kim, S., Razi, A., Stringhini, G., Wisniewski, P.J., & De Choudhury, M. (2021). A Human-Centered Systematic Literature Review of Cyberbullying Detection Algorithms. Proceedings of the ACM on Human-Computer Interaction, 5(CSCW2), Article 325. <https://doi.org/10.1145/3476066>
 9. Kumar, L., Goyal, P., Malik, K., & Kumar, R. (2020). Using Machine Learning, Image Processing & Neural Networks to Sense Bullying in K-12 Schools. *Asian Journal of Convergence in Technology*, 6(1). Retrieved from <http://www.asianssr.org>
 10. Konvencija o pravima deteta. (1989). Ujedinjene nacije.
 11. Lafrance St-Martin, L. I., & Villeneuve, S. (2024). The uses of chatbots in the context of children and teenagers bullying: a systematic literature review. *Cogent Education*, 11(1), 2312032. <https://doi.org/10.1080/2331186X.2024.2312032>

12. Milosavljević, M. (1998). Dijagnostikovanje osnovnih pojmova istraživanja. Zbornik Nasilje nad decom, Beograd: Fakultet političkih nauka, 9–46.
13. Milosevic, T., Van Royen, K., & Davis, B. (2022). Artificial Intelligence to Address Cyberbullying, Harassment and Abuse: New Directions in the Midst of Complexity. *International Journal of Bullying Prevention*, 4(1), 1–5. <https://doi.org/10.1007/s42380-022-00117-x>
14. Milosevic, T., Verma, K., Carter, M., Vigil, S., Laffan, D., Davis, B., & O'Higgins Norman, J. (2023). Effectiveness of Artificial Intelligence–Based Cyberbullying Interventions From Youth Perspective. *Social Media + Society*, 1–12. <https://doi.org/10.1177/20563051221147325>
15. Nansel, T. R., Overpeck, M., Pilla, R. S., Ruan, W. J., Simons-Morton, B., & Scheidt, P. (2001). Bullying behaviors among US youth: prevalence and association with psychosocial adjustment. *JAMA*, 285(16), 2094–2100. <https://doi.org/10.1001/jama.285.16.2094>
16. Olweus, D. (2013). School bullying: Development and some important challenges. *Annual Review of Clinical Psychology*, 9, 751–780.
17. Olweus, D. (1997). Bully/victim problems in school. *European Journal of Psychology of Education*, 7(4), 495–510.
18. Olweus, D. (2010). Understanding and researching bullying: Some critical issues. In *Handbook of bullying in schools: An international perspective* (pp.9–33). Routledge/Taylor & Francis Group: US.
19. Orrù, G., Galli, A., Gattulli, V., Gravina, M., Marrone, S., Micheletto, M., Procaccino, A., Nocerino, W., Terrone, G., Curtotti, D., Impedovo, D., Marcialis, G.L., & Sansone, C. (2023). Leveraging Artificial Intelligence to Fight (Cyber)Bullying for Human Well-being: The BullyBuster Project. *Proceedings of the Ital-IA 2023: 3rd National Conference on Artificial Intelligence*. Available at: <https://www.researchgate.net/publication/375006083>
20. Pham Van, T., & Pham, V.-N. (2023). UAV-Integrated IoT System for Handling Widespread Events. 2023 17th International Conference on Ubiquitous Information Management and Communication (IMCOM),

- Seoul, Korea, Republic of, 2023, pp. 1–8. <https://doi.org/10.1109/IMCOM56909.2023.10035602>
21. Popadić, D., Plut, D., & Pavlović, Z. (2014). *Nasilje u školama Srbije: Analiza stanja od 2006. do 2013. godine*. Beograd: Institut za psihologiju. UNICEF, Beograd.
22. Priručnik za primenu posebnog protokola za zaštitu dece i učenika od nasilja, zlostavljanja i zanemarivanja u obrazovno-vaspitnim ustanovama. (2009). Ministarstvo prosvete Republike Srbije.
23. Ptaszynski, M., Masui, F., Nitta, T., Hatakeyama, S., Kimura, Y., Rzepka, R., & Araki, K. (2016). Sustainable cyberbullying detection with category-maximized relevance of harmful phrases and double-filtered automatic optimization. *International Journal of Child-Computer Interaction*, 8. <https://doi.org/10.1016/j.ijcci.2016.07.002>
24. Raisi, E., & Huang, B. (2016). Cyberbullying Identification Using Participant-Vocabulary Consistency. *ArXiv*, abs/1606.08084.
25. Rajesh, Bahuguna., Minakshi, Memoria., & Rajiv, Kumar. (2023). Artificial Intelligence Based Legal Application For Combating Domestic Violence. 81–86. <https://doi.org/10.1109/InCACCT57535.2023.10141818>
26. Salmivalli, C., Lagerspetz, K., Björkqvist, K., Österman, K., & Kaukiainen, A. (1996). Bullying as a group process: Participant roles and their relations to social status within the group. *Aggressive Behavior: Official Journal of the International Society for Research on Aggression*, 22(1), 1–15.
27. Schoeler, T., Duncan, L., Cecil, C.M., Ploubidis, G.B., & Pingault, J.B. (2018). Quasi-experimental evidence on short- and long-term consequences of bullying victimization: A meta-analysis. *Psychol Bull.*, 144(12), 1229–1246. <https://doi.org/10.1037/bul0000171>
28. Silva, J. L., Oliveira, W. A., Mello, F. C. M., Andrade, L. S., Bazon, M. R., & Silva, M. A. I. (2018). Anti-bullying interventions in schools: a systematic literature review. *Ciência & Saúde Coletiva*, 22(7), 2329–2340. <https://doi.org/10.1590/1413-81232017227.16242015>

29. Soundar, K. R., & Ponesakki, P. (2016). Cyberbullying detection based on text representation. *International Journal of Engineering Science*, 6(10), 2776–2785.
30. Thomas, D., Warmesley, D., Macy, M., & Weber, I. (2017). Automated Hate Speech Detection and the Problem of Offensive Language. *International AAAI Conference on Web and Social Media*, 11(1), 512–515.
31. Topcu-Uzer, C., & Tanrikulu, I. (2018). Technological solutions for cyberbullying.
32. Yan, W., Yuan, Y., Yang, M., Zhang, P., & Peng, K. (2023). Detecting the risk of bullying victimization among adolescents: A large-scale machine learning approach. *Computers in Human Behavior*, 147. <https://doi.org/10.1016/j.chb.2023.107817>
33. ZAKON O OSNOVAMA SISTEMA OBRAZOVANJA I VASPITANJA (“Sl. glasnik RS”, br. 88/2017, 27/2018 – dr. zakon, 10/2019, 27/2018 – dr. zakon, 6/2020, 129/2021 i 92/2023).

PART III

ARTIFICIAL INTELLIGENCE AT YOUR SERVICE

THE IMPACT OF AI-SUPPORTED SERVICES ON CUSTOMER EXPERIENCE: A CASE STUDY OF THE HOTEL INDUSTRY IN SERBIA

Stevica DEĐANSKI¹

Faculty of Social Sciences, FNP, Belgrade, Serbia

Boris JEVTIĆ²

Faculty of Computer Science, RAF, Union University Belgrade, Serbia

Abstract: *Based on literature and current findings on the potential of digitalisation to enhance service quality models, this research aims to better understand the potential impact of service quality and personalization supported by artificial intelligence technologies on improving customer experience in the hotel industry. Responses from 218 participants from Serbia, who stayed at high-category hotels in Zlatibor, Vrnjačka Banja, and Kopaonik from May to July 2024, were collected. Using a five-point Likert scale and processed via regression analysis, these responses helped determine the significance of these effects. Results of this empirical research indicate that convenience, personalization, and digitalized hotel services leveraging AI technologies have a significant impact as exogenous factors in improving customer experience. This paper contributes to further analyses of new technologies and their potential to improve user interaction with services, as well as understanding the need for increased investments in AI, business modernization, service enhancement,*

1 dedjanskis@gmail.com

2 boris.jevtic10@gmail.com

operational capacity expansion, cost reduction, and development in the highly competitive tourism market.

Keywords: *Artificial Intelligence, service quality, experience economy, hotel industry.*

1. INTRODUCTION

Organizations face increasing customer demands, requiring shortened strategic planning periods and effective responses to market and technological pressures. Technological innovations, including those based on artificial intelligence, can assist businesses in the digital economy in accelerating the shift to digital ways of working, managing, organizing, and adapting to changes in organizational processes (Attaran, 2020; Rosario & Dias, 2022; Špiler et al., 2023; Miškić et al., 2024). Various theorists have explored aspects of service quality across different contexts, including Parasuraman, Zeithaml, and Berry (1994), Collier and Bienstock (2006), Scheidt and Chung (2019), and Suhartanto et al. (2019). However, they have rarely focused on the potential of AI-based shopping experiences, which could significantly alter perceptions of service quality. Understanding these changes enables organizations to adapt their commitment to customer relationships and evaluate the overall experience facilitated by AI, potentially leading to stronger connections between users and organizational services (Wang, Molina, and Sunder, 2020).

Implementing artificial intelligence (AI) has the potential to transform the way organizations in both private and public sectors interact with their clients (McLean & Osei-Frimpong, 2019). AI technologies, supported by data analytics, enable businesses to convert information about existing and potential user behavior into customer data, enhancing customer experience through increased insights into preferences and purchasing patterns (Evans, 2019). These technologies allow for service and recommendation personalization, guiding users towards the company's strategic objectives (Paschen et al., 2019). By analyzing past customer preferences (Maras, 2020), using AI-powered

chatbots, and examining customer insights at various key points of contact, organizations can automate their functions, reduce costs, increase flexibility, and streamline client interaction. These studies have inspired the authors to delve deeper into the possibilities of achieving a positive AI-based customer experience (Shank et al., 2019). This research focuses on areas of user interaction through modern AI technologies (Amin et al., 2020), used by service organizations to develop relationships with clients and improve their services.

The theoretical contributions include enhancing knowledge of human interaction with AI-supported services, while the practical implications of the research can assist service organizations in shaping investments and strategies for providing customer experiences through AI. Impact factors are defined based on the results of previous studies (de Medeiros et al., 2016), service quality models (Parasuraman et al., 1994), and user-related issues tied to AI-supported services (Davenport et al., 2020). As intermediary factors in the context of AI-based customer experience, personalization, convenience, and service quality in Serbia's hotel industry have been selected.

Hospitality is a vital and rapidly developing industry that operates in a highly competitive, service-based market and relies on customer reviews. Innovative technologies such as AI, if implemented by hotels, can place customer perspectives at the center of business operations. The global hotel industry is one of the most important and fastest-growing in the world (Ruel & Njoku, 2021). The use of AI technologies in hospitality services significantly impacts employee productivity. The quality of AI-based services contributes considerably to the overall service quality, necessitating a better understanding of how these technologies can affect customer satisfaction and loyalty, as well as employment, employee engagement, and service quality. Modern technological applications, including AI and robotics, alongside increased digital connectivity, affect all business sectors, including hospitality. Companies use these advanced technologies to improve operational processes, reduce costs, and enhance the customer experience (Mingotto et al., 2021). These technologies are applied in hotels, hospitality, and event management (Jiang, 2020; Berezina et al., 2019; Lu et al., 2020; Kumar et al., 2021; Thong-On et al., 2021).

The structure of the paper is organized so that after the abstract and introduction, the theoretical framework, research model, and scientific hypotheses follow. The empirical research is presented in the third section with key results, discussion, and conclusion. The scientific literature used is shown at the end of the paper.

2. THEORETICAL FRAMEWORK

This paper adopts the most widely used definition of artificial intelligence (AI) from the European Commission's Independent Expert Group, describing AI as the intelligent behavior of AI-based systems in analyzing the environment to achieve specific goals with a certain level of autonomy. These systems can be purely software-based, operating in virtual environments such as voice assistants, image analysis software, search engines, as well as speech and facial recognition systems. Alternatively, AI can also be integrated into hardware devices, including advanced robots, autonomous cars, drones, or devices connected to the Internet of Things. This definition provides a comprehensive overview of various forms and applications of artificial intelligence in the modern world (European Commission, Independent Expert Group, 2024). Previous research has highlighted the significant impact of artificial intelligence (AI) in the service sector, particularly in hospitality (Bisoï et al., 2020; Citak et al., 2021; Khatri, 2021; Srebro et al., 2024; Srebro & Jevtić, 2024). The paper defines the following research variables: customer experience, user personalization, and the AI-supported services themselves, which are further explained and analyzed.

2.1 Customer experience in AI-supported services

This experience can be analyzed as hedonistic, providing unforgettable, exciting, comfortable, entertaining, and educational experiences. Additionally, AI aids organizations in recognizing users and tailoring service design, enabling users to feel valued, respected, welcome, safe, and satisfied when using the service. Customer experience encompasses the entirety of a customer's

interaction with a seller, based on their interactions with and perceptions of the brand. Theories, models, and frameworks of customer experience allow for user-centered design activities, facilitate the choice of constructs and measures, and contextualize findings within a broader knowledge base. Customer experience includes users' perceptions and reactions resulting from actual or anticipated use of a product, system, or service (Davenport et al., 2020). According to Nielsen's definitions, customer experience represents a set of statements explaining observed phenomena, with a "model" used as a descriptive simplification of those phenomena and a "framework" as a lexicon for organizing constructs. In modeling customer experience, the authors focus on the following aspects:

- *User life cycle-oriented design development* – This includes a cognitive socio-technical framework and a framework of tasks, users, representations, and functions. These frameworks define user interactions in detail to facilitate the acquisition of customer experience at any stage of service development. The activity cycle, linking iterative design thinking with the software development life cycle for new technologies, includes AI. This process encompasses the pre-design phase, where customer experience elements are prepared during project planning, the design phase, where the organization defines user needs and designs system components, and the system integration and routine usage phase (Hartson & Pyla, 2012; Nielsen, 1994).
- *Cognitive socio-technical framework* – This framework integrates two complementary characterizations of user interactions: cognitive models, which explain how people process information, and socio-technical models, which show how technology affects work processes. This framework proposes evaluating interactions throughout the technology's life cycle. In the early stages, researchers study interactions between technology and individuals, while in later stages they analyze work processes within groups in context.

- *The framework of tasks, users, representations, and functions* – This socio-technical model consists of four steps, emphasizing interactions between users, tasks, and functions. The first step involves identifying all system users. The second step includes developing an ontology of activities, interactions, and knowledge to better understand work processes. The third step involves task analysis, where researchers catalog tasks related to complex activities to assess their complexity and duration. The fourth step relates to identifying and defining functions needed to support users in task completion, ensuring alignment between the system's technical and social aspects.

Based on previous studies, the following elements of customer experience can be analyzed: *cognitive* elements, such as perception, memory, language, problem-solving, and abstract thinking (American Psychological Association, 2016). According to Keiningham et al. (2017), these elements include service functionality, speed, and availability; *emotional* elements, which are complex in nature (Ladhari et al., 2017), encompassing positive or negative emotions like excitement, regret, anger, joy, or surprise during service use; *physical and sensory* elements, which distinguish offline from online customer experiences. Offline experiences are characterized by physical artifacts, lighting, appearance, and signage, while online experiences are associated with technology, including intuitive user interfaces and clear design; and *social* elements, which relate to the influence of others, including family, friends, and broader social networks. These elements also encompass users' social identity, referring to their self-perception and how they are perceived in a social context.

Since AI technologies, such as machine learning, natural language understanding, and natural language processing, can aid in analyzing user sentiment and gathering feedback with precision, speed, and scale that is difficult to achieve solely through human engagement (Gartner, 2020), AI has the potential to become a crucial tool for service providers in continuously

enhancing customer experience, helping them remain competitive in the market (Newman, 2019). Therefore, AI-supported customer experience in hospitality is defined as the dependent variable in this research. In particular, AI in hospitality and tourism is increasingly used alongside augmented reality, image recognition through computer vision, and predictive inventory (Saponaro et al., 2018). For these technologies to effectively enhance customer experience, a detailed understanding of users through data and client profiles, including their preferences and past experiences, is essential.

2.2 Service convenience and personalization through AI technologies

These factors are key areas in the modern relationship between service organizations and customers, requiring significant investments. However, these investments are still insufficiently prioritized in overall technological investments, where a focus typically lies on technologies that enhance business processes and the service itself, rather than communication and these aspects. Service convenience is achieved by saving time and effort and enabling mobility, which can be crucial for attracting service-interested customers. The time-saving aspect of convenience has been studied through the impact of perceived waiting time on customer experience. AI-powered chatbots can proactively initiate conversations with users, provide relevant information, and assist at every point of contact throughout the customer lifecycle, allowing users to get answers when they need them without waiting for an available employee, which improves resolution time and customer satisfaction (Walch, 2019). Convenience contributes to a positive customer experience by reducing or eliminating barriers for users, thereby increasing the trust that users have in the service and the technology used to deliver it. Additionally, the perception of convenience influences users' overall evaluation of the service's utility. Enhancing service personalization and customer experience through AI is particularly important in the modern

communication mix of service organizations. Personalization involves tailoring information to the specific needs of each user, leading to positive experiences (Zhang, Edwards, & Harding, 2007). Data mining techniques enable a higher level of purchasing interest, and personalization in online services can be observed through the following dimensions: personalization of the user interface, which includes adaptability of screen appearance and overall presentation; content personalization, which involves differentiating information based on individual user profiles, including product or service offerings and prices; and personalization of the interaction process, where AI algorithms autonomously decide when and how to communicate with users. Personalization strategies that generate highly positive reactions from users strengthen their commitment to the service and organization. Based on these insights, perceived time and spatial convenience of services, interface personalization, content, and interactions – all through AI technologies – can have positive implications for the customer experience.

1.3 Quality of services supported by AI technologies

This includes interface design, customer support, security, and reliability. Since many AI-supported services are based on a self-service model, a well-designed user interface is often considered a key success factor for these services. Artificial intelligence can transform the user interface by controlling all design aspects, including visual elements, typography, animations, and graphical information (Irfan, 2020). Previous research confirms that technical and functional service quality influences how customers evaluate brands. Studies on self-service technologies suggest that users most commonly assess service quality based on security, reliability, customer support, and interface design dimensions (McKecnie, Ganguli & Roi, 2011; Wolfinbarger & Gilli, 2003). The quality of AI-based services largely depends on the quantity and quality of

personal information that the service provider can collect about users. Although most of this data is not highly sensitive, the combination of artificial intelligence and seemingly insignificant data, such as marketing choices and preferences, can lead to the creation of extensive user profiles, making their protection more challenging. Nevertheless, research (Saratchandran, 2019) suggests that artificial intelligence can significantly enhance the reliability of customer services. Chatbots and other AI-based customer support tools are increasingly used as an automated and efficient means of improving customer services (Treasure Data, 2019; Curčić, Grubor, and Jevtić, 2024). Previous studies also confirm the impact of service quality on perceived value, which relates to the balance between the benefits users receive and the sacrifices they must make in exchange for the service. The challenge of providing hospitality while maintaining social distancing can be effectively addressed through AI-based interactions. These interactions include service encounters that are supported, mediated, or fully generated by AI technologies. Traditionally, service encounters are viewed as a form of social exchange between the user and the service provider. Based on academic literature and practice, numerous AI applications in customer encounters have been identified across various service contexts, using a technology-context matrix.

3. METHODOLOGY

1.1 Research model

Based on the theoretical framework, a research model was defined (Figure 1), comprising two independent variables:

- F_1 = AI-supported service personalization and convenience
- F_2 = Quality of services supported by artificial intelligence (AI)

and one dependent variable:

- F_3 = Customer experience supported by artificial intelligence (AI)

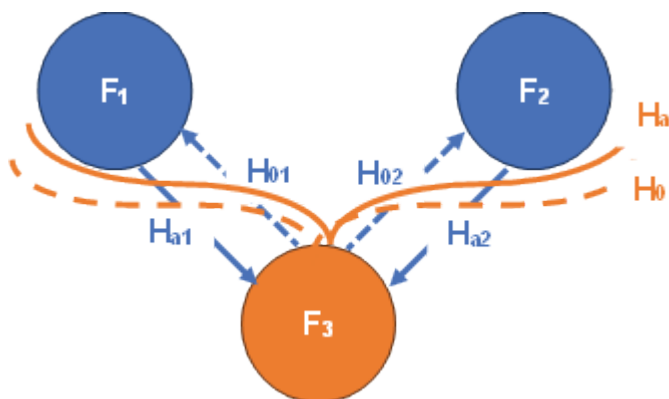


Figure 1. Research model

Source: Authors

The research objectives are:

- To determine whether F_1 significantly impacts F_3 .
- To determine whether F_2 significantly impacts F_3 .

The ultimate objective is to determine whether both F_1 and F_2 significantly impact F_3 .

Main hypotheses:

- H_0 : F_1 and F_2 do not impact F_3
- H_a : F_1 and F_2 impact F_3

Sub-hypotheses:

- H_{01} : F_1 does not impact F_3
- H_{a1} : F_1 impacts F_3
- H_{02} : F_2 does not impact F_3
- H_{a2} : F_2 impacts F_3

1.2 Sample for empirical research

For this research, an online survey was used to collect responses from 218 participants from Serbia who stayed in five- and four-star hotels in Zlatibor, Vrnjačka Banja, and Kopaonik between May and July 2024.

A case study of the hotel industry in Serbia

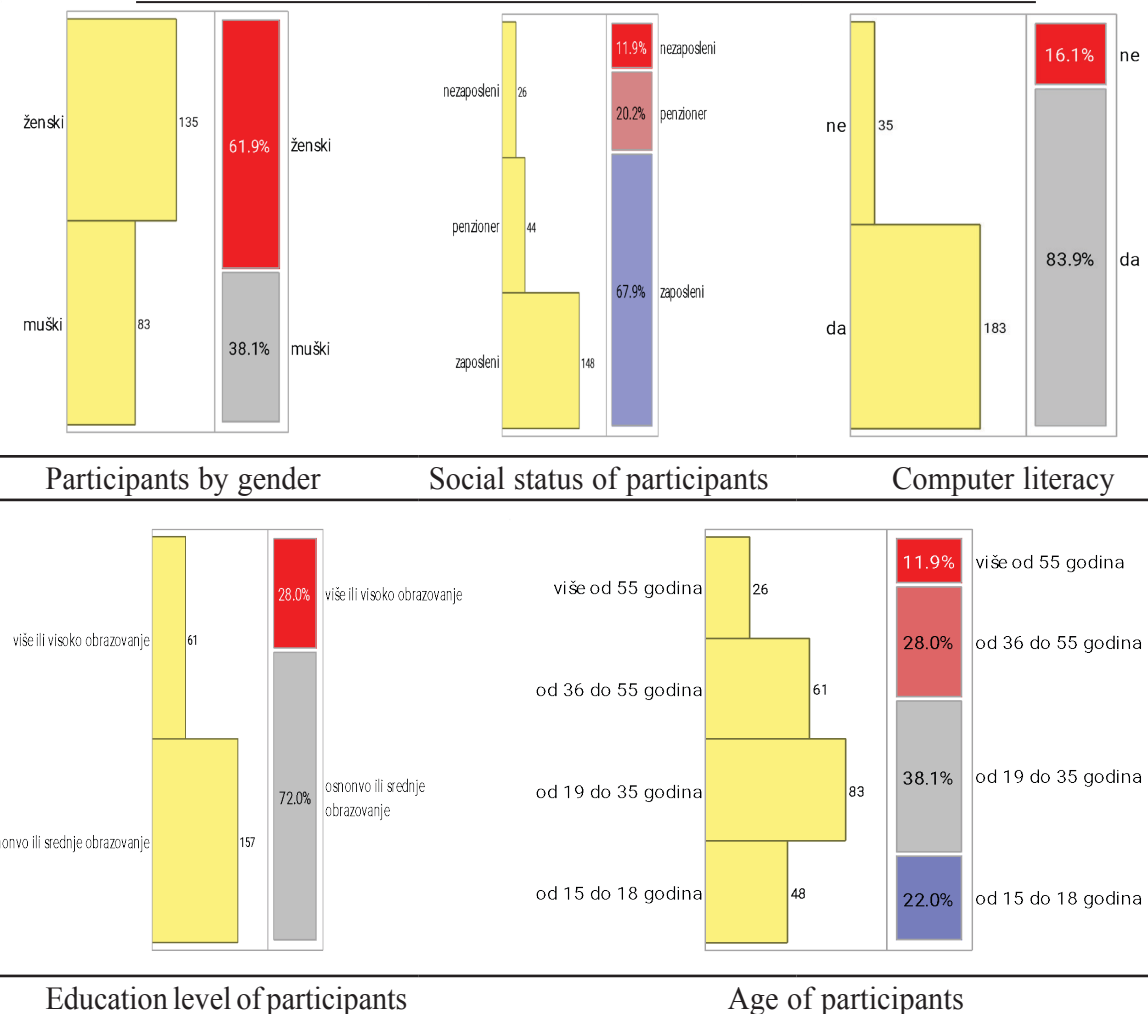


Figure 2: Frequency and probability of participant characteristics

Participants responded based on their hotel experience concerning the quality of hotel services, the level of service personalization, and the convenience of AI-supported hotel services, assessing the significance of these modernized services on their experience. Correlation and regression analysis was used, and processed in SAS JMP Pro 18 statistical software. The survey used a questionnaire where participants rated statements related to the research model variables F1, F2, and F3 on a five-point Likert scale, with 5 as the highest score.

The frequency and probability of participant characteristics by gender, age, education level, computer literacy, and social status are shown in Figure 2. Of the total 218 participants, 135 were women (61.92%), and 83 were men (38.07%). The majority were between 18 and 35 years old (83 participants, 38.07%), with the least number over 55 years old (26 participants, 11.92%). There were more participants with completed primary or secondary education (157, 72.01%), and 61 (27.98%) had higher education. A total of 183 participants (83.94%) had basic computer literacy, while 35 (16.05%) did not. In terms of social status, most participants were employed (148, 67.89%), with the least number unemployed (26, 11.92%) out of the total 218 participants.

1.3 Results

Table 1 shows the mean and standard deviation for the statements (3 x 3 = 9) related to the research variables F₁, F₂, and F₃.

Table 1: Results for research variables

Statement	Mean	Standard Deviation
F1		
F1a	3.7201834862	1.0158449586
F1b	4.0275229358	0.874194461
F1c	3.9449541284	1.0545915617
F2		
F2a	3.9311926606	1.1559660709
F2b	3.8532110092	1.0675411092
F2c	3.7247706422	1.0193660236
F3		
F3a	4.0596330275	0.9113335816
F3b	3.7752293578	1.0204955843
F3c	3.9633027523	1.0939852778

Table 3: Mean and standard deviation for variables

Variable	F1	F2	F3
Mean	3.8975535168	3.8363914373	3.9327217125
Standard Deviation	0.8961453143	0.7277970222	0.6701595388

Variable F_3 has the highest mean response value of 3.93, while variable F_1 has the highest standard deviation of 0.90.

Correlation analysis

Pearson correlation in Table 4 shows the highest correlation coefficient between variables F_1 and F_2 at 0.8574, indicating that independent variable F_2 can be explained by independent variable F_1 by 73.51%, and the lowest is between variables F_1 and F_3 at 0.7677, meaning dependent variable F_3 can be explained by independent variable F_1 by 58.94%. A positive correlation was found among all research variables.

Table 4. Correlation model

	F1	F2	F3
F1	1.0000	0.8574	0.7677
F2	0.8574	1.0000	0.8329
F3	0.7677	0.8329	1.0000

The statistical significance of F_1 on F_3 is shown in Table 5, with the ANOVA test confirming the null hypothesis that $r^2 = 0$, with a significance level below 0.0001 [$F(1,216) = 310.0897$, $p < 0.0001$]. This confirms a significant impact of F_1 , related to AI-supported service convenience and personalization, on F_3 , i.e., customer experience in hotel services.

Table 5. F1 and F3 ANOVA

Source	DF	Sum of squares	Mean Square	F ratio
Model	1	57.443869	57.4439	310.0897
Error	216	40.013827	0.1852	Probability > F
Total	217	97.457696		< 0.0001

Table 6 presents the contribution values of F_1 in predicting dependent variable F_3 at 0.767739, meaning 58.94% of F_3 can be explained by F_1 . The VIF level is 1.000, confirming the alternative hypothesis H_{a1} that F_1 impacts F_3 as valid.

Table 6. F1 and F3 coefficients

Variable	Value	Standard error	t ratio	Probability > t	Standard Beta coefficients	VIF
Intercept	1.6950023	0.130376	13.00	< 0.0001	0	.
F_1	0.5741344	0.032604	17.61	< 0.0001	0.767739	1

Regression equation (1) is derived based on data from the table above:

$$F3 = 1.6950023 + 0.5741344 \cdot F1$$

$$F3 = 1.6950023 + 0.5741344 \cdot F1$$

(1)

Further calculations in Table 7 and ANOVA test confirm the null hypothesis $r^2 = 0$, with a significance level below 0.0001 [$F(1,216) = 489.3207$, $p < 0.0001$], indicating a significant effect of F_2 on F_3 .

Table 7. F2 and F3 ANOVA

Source	DF	Sum of squares	Mean square	F ratio
Model	1	67,611895	67,6119	489,3207
Error	216	29,845801	0,1382	Probability > F
Total	217	97,457696		< 0,0001

Table 8 shows the impact values of F_2 on the projection of F_3 at 0.83292, meaning 69.38% of F_3 can be explained by F_2 . VIF is 1.000, and the alternative hypothesis H_{a2} that F_2 impacts F_3 is accepted.

Table 8. Coefficients for variables F1 and F3

Variable	Value	Standard error	t ratio	Probability > t	Standardized Beta coefficients	VIF
Intercept	0.9903718	0.135376	7.32	< 0.0001	0	.
F_2	0.7669577	0.034672	22.12	< 0.0001	0.83292	1

The regression equation (2), based on the data in Table 8, is:

$$F3 = 0.9903718 + 0.7669577 \cdot F2$$

$$F3 = 0.9903718 + 0.7669577 \cdot F2$$

(2)

The independent variables are F_1 and F_2 , and the dependent variable is F_3 . The statistical significance is presented in Table 9, and the ANOVA test confirms the null hypothesis that $r^2 = 0$, as the significance is less than 0.0001 [$F(2,215) = 256.4180$, $p < 0.0001$].

Table 9. F_1 , F_2 and F_3 ANOVA

Source	DF	Sum of squares	Mean square	F ratio
Model	2	68.669063	34.3345	256.4180
Error	215	28.788633	0.1339	Probability > F
Total	217	97.457696		< 0.0001

Table 10 provides the multiple contributions of the independent variables F_1 and F_2 in predicting the dependent variable F_3 . The independent variable F_2 has a greater contribution to the dependent variable F_3 (0.659416), while F_1 has a smaller contribution (0.202364). The total combined contribution is 70.46%. The multiple correlation coefficient is 0.8394, which is highly positive. The VIF level is 1.000, confirming the alternative hypothesis H_a that F_1 and F_2 impact F_3 as valid.

Table 10. Coefficients for variables F_1 , F_2 and F_3

Variable	Value	Standard error	t ratio	Probability > t	Standardized Beta coefficients	VIF
Intercept	1.0134608	0.133518	7.59	< 0.0001	0	.
F_1	0.1513331	0.053858	2.81	0.0054	0.202364	3.7752061
F_2	0.6071935	0.066316	9.16	< 0.0001	0.659416	3.7752061

The multiple linear regression equation (3), derived from Table 10 is:

$$F3 = 1.0134608 + 0.1513331 \cdot F1 + 0.6071935 \cdot F2$$

$$F3 = 1.0134608 + 0.1513331 \cdot F1 + 0.6071935 \cdot F2$$

(3)

All alternative hypotheses are accepted – (H_{a1} : F_1 impacts F_3), (H_{a2} : F_2 impacts F_3) i (H_a : F_1 and F_2 impact F_3).

4. CONCLUSION

The use of artificial intelligence is a strategic and critical factor in the economic development of the service industry, especially hospitality, which already invests significantly in new technologies and business modernization, service digitization, and customer communication. AI technologies enabling the use of large amounts of data are increasingly becoming digital assistants, helping service enterprises improve services for customers, expand operational capabilities, reduce costs, enhance customer experience, and retain tourists by offering pleasant, longer-term interactions with them (Khatri, 2021; Jevtić & Dedjanski, 2013).

Results from this study, using data from tourists in Serbian hotels during the first half of 2024, demonstrate a high level of interest in innovative technologies through the perception of AI experiences. New methods of service user interaction, communication, personalization, and digitalization are increasingly important endogenous factors that hoteliers should consider when introducing AI technologies. Study findings indicate that these elements of services play a significant role in users' experiences with AI. The complexity of AI technology from the user perspective, gaining their interest in services before use, trust during their hotel stay, and the influence of their views shared on social media post-usage pose a considerable challenge for service organizations. These organizations must invest in innovations to effectively apply AI technologies and enhance customer experience, which is a crucial exogenous factor for development.

From the perspective of future research, the results of this study, focused on the individual's customer experience in the context of AI-supported services, encourage further examination of other endogenous factors that may provide a more nuanced perspective on the success factors of AI-supported services, as well as on demographic customer segments (Jevtić et al., 2024).

LITERATURE

1. Ameen, N., Tarhini, A., Shah, M., & Hosany, S. (2020). *Consumer interaction with cutting-edge technologies*. <https://www.journals.elsevier.com/computers-in-human-behavior/call-for-papers/consumer-interaction-with-cutting-edge-technologies>. (Pristupljeno: 2. 6. 2024).
2. American Psychological Association. (2016). *Glossary of psychological terms*. <https://www.apa.org/research/action/glossary.aspx>. (Pristupljeno: 2. 6. 2024).
3. Attaran, M. (2020). Digital Technology Enablers and Their Implications for Supply Chain Management. *Supply Chain Forum: An International Journal*, 21(3), 158–172.
4. Berezina, K., Ciftci, O., & Cobanoglu, C. (2019). Robots, Artificial Intelligence, and Service Automation in Restaurants. In *Robots, Artificial Intelligence, and Service Automation in Travel, Tourism, and Hospitality*, Emerald Publishing Limited, Bingley, 185–219.
5. Collier, J. E., & Bienstock, C. C. (2006). Measuring service quality in E-retailing. *Journal of Service Research*, 8(3), 260–275.
6. Čurčić, N., Grubor, A., & Jevtić, B. (2024). Implementing Artificial Intelligence in Travel Services: Customer Satisfaction Gap Study at Serbian Airports. *Ekonomika*, 3/2024.
7. Davenport, T., Guha, A., Grewal, D., & Bressgott, T. (2020). How artificial intelligence will change the future of marketing. *Journal of the Academy of Marketing Science*, 48(1), 24–42.
8. Evans, M. (2019). *Build A 5-star customer experience with artificial intelligence*. <https://www.forbes.com/sites/allbusiness/2019/02/17/customer-experience-artificial-intelligence/#1a30ebd415bd>. (Pristupljeno: 6. 5. 2024).
9. Gartner. (2020). *Drive growth in times of disruption*. <https://www.gartner.com/en>. (Pristupljeno: 8. 6. 2024).
10. Irfan, M. (2020). *Artificial intelligence and the future of web design*, 22 February 2020. <https://usabilitygeek.com/>

- [artificial-intelligence-and-the-future-of-web-design/](#). (Pristupljeno: 10. 6. 2024).
11. Jevtić, B., Dedanski, S., Beslać, M., Grozdanić, R., & Damnjanović, A. (2013). SME Technology Capacity Building for Competitiveness and Export - Evidence from Balkan Countries. *Metalurgija International*, 18(spec. iss. 4), 162–170. București: Editura Științifică F.M.R. <https://enauka.gov.rs/handle/123456789/676927>.
 12. Jevtić, B., Beslać, M., Janjušić, D., & Jevtić, M. (2024). The Effects of Digital Natives' Expectations of Tech Hotel Services Quality on Customer Satisfaction. *International Journal for Quality Research*, 18(1), 1–10. DOI: 10.24874/IJQR18.01-01. www.ijqr.net/paper.php?id=1134
 13. Khatri, M. (2021). How Digital Marketing along with Artificial Intelligence is Transforming Consumer Behaviour? *International Journal for Research in Applied Science & Engineering Technology*, 9(7), 523–527.
 14. Keiningham, T., Ball, J., Benoit, S., Bruce, H. L., Buoye, A., & Dzenkovska, J. et al. (2017). The interplay of customer experience and commitment. *Journal of Services Marketing*, 31(2), 148–160.
 15. Kumar, S., Kumar, V., & Attri, K. (2021). Impact of Artificial Intelligence and Service Robots in Tourism and Hospitality Sector: Current Use & Future Trends. *Administrative Development: A Journal of HIPA, Shimla*, 8(SI-1), 59–83.
 16. Ladhari, R., Souiden, N., & Dufour, B. (2017). The role of emotions in utilitarian service settings: The effects of emotional satisfaction on product perception and behavioral intentions. *Journal of Retailing and Consumer Services*, 34, 10–18.
 17. Lu, V. N., Wirtz, J., Kunz, W. H., Paluch, S., Gruber, T., Martins, A., & Patterson, P. G. (2020). Service Robots, Customers, and Service Employees: What Can We Learn from the Academic Literature and Where Are the Gaps? *Journal of Service Theory and Practice*, 30(3), 361–391.

18. McLean, G., & Osei-Frimpong, K. (2019). Hey Alexa... examine the variables influencing the use of artificial intelligent in-home voice assistants. *Computers in Human Behavior*, 99, 28–37.
19. Wolfenbarger, M., & Gilly, M. C. (2003). eTailQ: Dimensionalizing, Measuring and Predicting eTail Quality. *Journal of Retailing*, 79(3), 183–198.
20. de Medeiros, J. F., Ribeiro, J. L. D., & Cortimiglia, M. N. (2016). Influence of perceived value on purchasing decisions of green products in Brazil. *Journal of Cleaner Production*, 110, 158–169.
21. McKecnie, S., Ganguli, S., & Roy, S. K. (2011). Generic technology-based service quality dimensions in banking. *International Journal of Bank Marketing*, 29(2), 168–189.
22. Mingotto, E., Montaguti, F., & Tamma, M. (2021). Challenges in Re-Designing Operations and Jobs to Embody AI and Robotics in Services: Findings from a Case in the Hospitality Industry. *Electronic Markets*, 31(3), 493–510.
23. Miškić, M., Srebro, B., Rašković, M., & Vrbanac, M., Jevtić, B. (2024). Key Challenges Hindering SMEs' full benefit from Digitalization – A Case Study from Serbia. *International Journal for Quality Research*, 19(2). DOI: 10.22874/IJQR1902-03.
24. Newman, D. (2019). *5 ways AI is transforming the customer experience*. <https://www.forbes.com/sites/danielnewman/2019/04/16/5-ways-ai-is-transforming-the-customer-experience/#49e17a31465a>. (Pristupljeno: 7. 7. 2024).
25. Nielsen, J. (1994). *Usability engineering*. Elsevier.
26. Paschen, J., Kietzmann, J., & Kietzmann, T. C. (2019). Artificial intelligence (AI) and its implications for market knowledge in B2B marketing. *Journal of Business & Industrial Marketing*, 34(7), 1410–1419.
27. Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1994). Reassessment of expectations as a comparison standard in measuring service quality: Implications for further research. *Journal of Marketing*, 58(1), 111–124.

28. Rosário, A. T., & Dias, J. C. (2022). Sustainability and the Digital Transition: A Literature Review. *Sustainability*, 14, 4072.
29. Ruel, H., & Njoku, E. (2021). AI Redefining the Hospitality Industry. *Journal of Tourism Futures*, 7(1), 53–66.
30. Saponaro, M., Le Gal, D., Gao, M., Guisiano, M., & Maniere, I. C. (2018). Challenges and Opportunities of Artificial Intelligence in the Fashion World. In *2018 International Conference on Intelligent and Innovative Computing Applications (ICONIC)*, 1–5. IEEE.
31. Scheidt, S., & Chung, Q. B. (2019). Making a Case for Speech Analytics to Improve Customer Service Quality: Vision, Implementation, and Evaluation. *International Journal of Information Management*, 45, 223–232.
32. Suhartanto, D., Helmi Ali, M., Tan, K. H., Sjahroeddin, F., & Kusdibyo, L. (2019). Loyalty Toward Online Food Delivery Service: The Role of E-Service Quality and Food Quality. *Journal of Foodservice Business Research*, 22(1), 81–97.
33. Shank, D. B., Graves, C., Gott, A., Gamez, P., & Rodriguez, S. (2019). Feeling Our Way to Machine Minds: People's Emotions When Perceiving Mind in Artificial Intelligence. *Computers in Human Behavior*, 98, 256–266.
34. Saratchandran, V. (2019). *Artificial Intelligence (AI): Ways AI is Redefining the Future of Customer Service*. <https://becominghuman.ai/artificial-intelligence-ai-ways-ai-is-redefining-the-future-of-customer-service-4dc667bfa59>. (Pristupljeno: 11. 6. 2024).
35. Srebro, B., & Jevtić, B. (2024). Improving Decision-Making Efficiency Through AI-Powered Fraud Detection and Prevention. In *International Congress on Project Management of ICT*, Arandelovac, 2024.
36. Srebro, B., Paunović, L., & Jevtić, B. (2024). Unraveling Hospitality: Exploring Human, Digital, and External Forces in Marketing Communications. In *Symorg*, 2024.
37. Špiler, M., Milošević, D., Miškić, M., Gostimirović, L., & Beslać, M. Jevtić, B. (2023). Investments in Digital Technology Advances in Textiles.

- Industria Textila*, 74(1), 90–97. DOI: 10.35530/IT.074.01.202287. <https://enauka.gov.rs/handle/123456789/811223>.
38. Treasure Data. (2019). *AI vs. Human Customer Service: Survey Data Shows When Consumers Prefer a Bot*. <https://app.hushly.com/runtime/content/hweB8PE1UYgB6oQR>. (Pristupljeno: 22. 5. 2024).
39. Tong-On, P., Siripipatthanakul, S., & Phayaphrom, B. (2021). The Implementation of Business Intelligence Using Data Analytics and Its Effects Towards Performance in Hotel Industry in Thailand. *International Journal of Behavioral Analytics*, 1(2), 1–16.
40. Walch, K. (2019). *AI's Increasing Role in Customer Service*. <https://www.forbes.com/sites/cognitiveworld/2019/07/02/ais-increasing-role-in-customer-service/#4ffaba4c73fc>. (Pristupljeno: 2. 8. 2024).
41. Wang, J., Molina, M. D., & Sundar, S. S. (2020). When Expert Recommendation Contradicts Peer Opinion: Relative Social Influence of Valence, Group Identity, and Artificial Intelligence. *Computers in Human Behavior*, 107, 106278.

DIGITAL TECHNOLOGIES, AI AND CONSUMER BEHAVIOR: THE INFLUENCE OF INFLUENCERS

Svetlana Kralj¹

Independent Researcher; Podgorica, Montenegro

Tatjana Mamula Nikolić²

Metropolitan University; Belgrade, Serbia

Abstract: *The goal of the paper is to conduct a comprehensive secondary research study to analyze the interplay between digital technologies, artificial intelligence (AI), influencers, and consumer behavior. Through the synthesis of existing literature and data from relevant sources, the study examines how influencers leverage digital platforms like YouTube, Instagram, and TikTok to influence consumer preferences and purchasing decisions. The COVID-19 pandemic has accelerated the reliance on influencers as consumers increasingly shift to online shopping and trust recommendations from their favorite digital personalities. Results show that AI tools play a pivotal role in enhancing influencer effectiveness by enabling personalized content creation and leveraging data-driven insights to optimize engagement and conversion rates for brands. Additionally, the emergence of AI-generated virtual influencers marks a notable trend in expanding the influence landscape. Secondary research underscores the transformative impact of digital technologies and AI on reshaping consumer*

1 ana@connection-s.com (Svetlana Kralj)

2 tatjana.mamula@metropolitan.ac.rs

behavior and marketing strategies. To conclude, it could be said that brands can achieve higher returns on investment (ROI) by harnessing the synergy between influencers and AI-driven analytics. This study provides a robust understanding of how these elements intersect, offering valuable insights into navigating the evolving landscape of digital marketing.

Keywords: *digital technologies, AI, influencers, influencer marketing, consumer behavior, social media marketing, virtual influencers, generational cohorts, personalized content, data-driven insights.*

INTRODUCTION

In modern marketing, particularly social media marketing, influencers play an exceptional role. Although the concept of influencers is not new, evolving consumer needs and behaviors, especially during the pandemic, have increased their significance. In the conditions in which companies were faced with labor shortages, poor logistics, and global economic changes, it is logical that they increasingly turned to influencers, giving them an advantage over traditional marketing tools.

On the other side, consumers are increasingly opting for online shopping and often opting for what they have seen on various social networks. On these platforms, consumers follow influencers who attract their attention with their knowledge, skills, beauty, photos, videos, and similar attributes. Through following, they build relationships based on trust and affinity.

With the emergence of TikTok as a new social application, with over 1 billion active users worldwide, companies began to build their strategies based on influencer marketing. In the era of high competition, gaining attention on social networks has become increasingly difficult, especially for small and medium-sized companies. Many companies struggled to adapt to this approach.

216 Consequently, it is not surprising that companies are setting a budget for

influencer marketing, and according to the Influencer Marketing Hub report, that number is expected to reach \$24 billion by the end of 2024.

Supporting the notion that influencer marketing is becoming extremely popular as a key marketing channel that marketers can use as part of their strategies, a 2023 HubSpot report indicates that 25% of marketers collaborate with influencers to promote products and services, making it a more popular marketing channel than virtual events, webinars and social commerce (HubSpot, 2023).

Global influencer market size from 2020 to 2025
(in billion U.S. dollars)

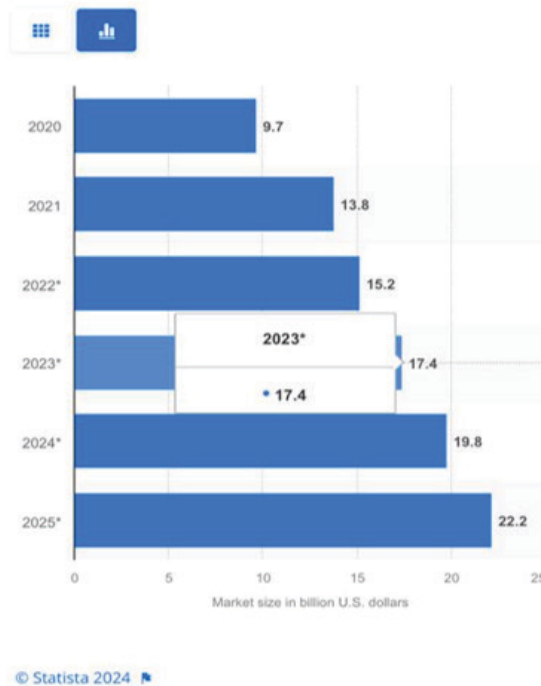


Image 1: Influencer marketing market size worldwide from 2016 to 2024 (in billion U.S. dollars), available at <https://www.statista.com/statistics/1092819/global-influencer-market-size/>.

According to Statista (Image 1), the global market value of influencer marketing was 21.1 billion US dollars in 2023, more than tripling compared to

2019. Influencers promote products and services for every industry and are eager to partner with companies. As influencer marketing continues to evolve as an industry, collaboration between brands and creators is becoming more profitable than ever.

Marketers see the main role of influencers in their audience. Influencer audiences provide value to marketers by offering organic reach, specific targeting, and increased attention. Brands leverage the power of sponsoring influencers to promote their content on their behalf to reach more and more of their target audience. Effective influencer collaborations will usually have a niche audience that is likely to have targeted appeal to maximize content engagement.

An influencer is a respected figure on social media who is compensated by companies, either monetarily or through free products, services, trips, and so on. Influencers are not only famous personalities but also individuals from various walks of life, who have many followers and whose recommendations are trusted. They create content to promote the products or services of the brand they collaborate with. Influencers almost always work on social media. They use platforms such as TikTok, Instagram, Facebook, YouTube, Pinterest, Twitch, LinkedIn and others. The primary role of influencers is to influence consumer behavior.

The task of influencers is to build a relationship with the audience by sharing their daily life, knowledge, and information and talking to other users on social networks. With significant follower bases, influencers can act as effective marketing agents, consistently posting valuable content on social media in support of the brand (Campbell and Farrell, 2020).

Today, many influencers are present on the market thanks to the effective results of influencer marketing, as well as the increased number of social network users who want to cooperate with brands for promotion, receiving compensation, or free products and services in return (Govindan and Alotaibi, 2021).

Influencers are often divided into bloggers, YouTubers, Instagrammers, TikTokers, celebrities, influencers, gaming influencers, and virtual or AI influencers. According to the number of followers, they are classified as (Campbell and Farrell, 2020):

Nano Influencers – 1–10k followers.

Micro-Influencers – 10–100k followers.

Macro Influencers – 100k–1 million followers.

Mega Influencers – 1 million+ followers.

Influencers choose platforms based on their number of followers and the industry of the products or services they promote. Influencer marketing has become an essential digital marketing strategy component for marketing and profit production. Industry data shows an upsurge in the use of influencer agencies.

1. AI AND INFLUENCERS

With the digital revolution comes the expansion of new technologies, primarily AI technology. It is safe to say that AI has entered every aspect of life and is everywhere around us. Consequently, the rise of AI tools and AI influencers is unsurprising.

According to the HubSpot report (2024) *“The 2024 State of Marketing & Trends Report: Data from 1400+ Global Marketers”* (Image 2), brands have readily welcomed new technologies, and are now trying new media formats using AI tools. The report further states that 85% of marketers say generative AI has changed how they will create content in 2024 and 63% predict that most content in 2024 will be created at least in part with the help of generative AI (Hubspot, 2023).

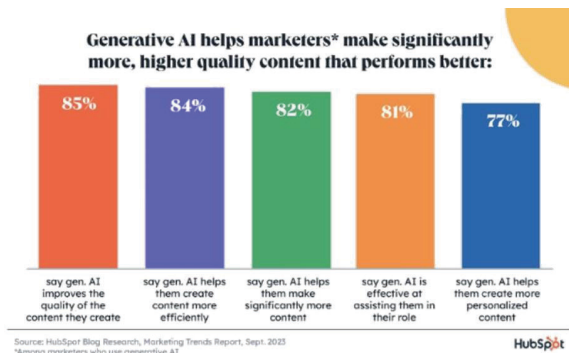


Image 2: <https://blog.hubspot.com/marketing/hubspot-blog-marketing-industry-trends-report>

Most people are already familiar with AI chatbots and image generators, which use them to create convincing images and text at incredible speed. Tools such as GPT-4, DALL-E 2, and Mid-journey Stable Diffusion have the potential to drastically change how advertisers reach and engage with customers. The potential applications of these powerful tools signal a new era in the marketing and advertising industries, providing unprecedented chances for growth and optimization (Yoo and Piscarac, 2023). This is the great power of generative AI or genAI: it uses algorithms to generate new content – writing, images, or audio – from training data. The application of genAI in marketing is multifaceted: from marketing strategies, through the analysis of consumer behavior to sending personalized messages, feedback and the like, which results in higher engagement and conversion rates (Deveau et al., 2023).

According to HubSpot's report, social media are becoming the most popular channels, and short video formats such as TikTok, YouTube shorts, and Instagram Reels dominate the use of AI tools. The most common content that uses generative AI includes social media posts, emails, blogs, topic ideas, images, and product descriptions. Using AI tools in influencer marketing does not mean that the tools will do all the work for them. With the help of ChatGPT, influencers produce content that is consistent with the voice of the brand they represent more easily. ChatGPT helps influencers determine the best time to post so that their post has greater visibility and engagement (Saini et al., 2024).

In addition to using generative AI to create content, influencers also use AI tools such as Synesthesia or DeepBrain to edit video content quickly and easily. They also use AI tools to respond to emails and messages to help with carrying out engagement activities. The DreamGF platform digitizes influencers and creates AI versions of them that can chat as if they were real people. Generative AI has many applications in assisting with the day-to-day work of an influencer (Marr, 2023). The number of AI tools is increasing daily, thus their use and testing by influencers.

While influencers use their authenticity and creativity to engage their audience, AI tools enhance their performance and engagement. On the other hand, the emergence of AI influencers, such as virtual influencers, brings new dimensions and challenges to digital marketing. AI-generated influencers, or virtual influencers, are on the rise in digital marketing. They are designed to look and act like real people, and have personality traits. They provide brands with more control and authority over sponsored content and are a more cost-effective alternative to social media influencers (Deveau et al., 2023). The key benefit for corporations adopting AI influencers is increased control over branding and messaging based on firm preferences. AI influencers are more effective in engaging clients since they can conduct meaningful discussions and learn from each interaction through AI assessment of behavior and reactions to various variants (Gerlich, 2023).



Lil Miquela – [@lilmiquela](https://www.instagram.com/lilmiquela)

Image 3: AI-Generated Avatar Influencers (Conversation, 2023). <https://theconversation.com/virtual-influencers-meet-the-ai-generated-figures-posing-as-your-new-online-friends-as-they-try-to-sell-you-stuff-212001>

The first AI creator, Lil Miquela, was created in 2016 by the startup Brud (Image 3). Miquela, a computer-generated 19-year-old female, has 2.6 million

Instagram followers and has partnered with some of the largest brands, including Prada, Calvin Klein, BMW, and Samsung. Miquela is also a recognized Spotify creator, having produced a collection of her music and appeared in music videos. After breaking up with the Kardashians, the French brand Balmain created three AI influencers: Shuda, Margot, and Ji, who, according to their statement, were a better reflection of the brand's inclusion. Global brands, including KFC, LVMH, Mini, Netflix, Nike, and Samsung, have partnered with AI influencers (Sands et al., 2022).

According to the Influencer Marketing Hub Report, the results of a survey conducted on more than 3,000 marketing agencies, brands, and other relevant professionals about the current state of influencer marketing asked respondents if they plan to use artificial intelligence and ML over the next year to identify influencers or create effective campaigns. A very substantial 63% of respondents said they would, along with 27.1% who thought they would. Only 9.9% gave an unequivocal no to this question. More than 60% of respondents have used virtual influencers (Influencer Marketing Hub, 2024).

Data from HubSpot's survey of 1,460 B2B and B2C marketers worldwide to reveal the top marketing trends in 2024 says that with image generators, content generators, and video generators, among other tools, AI is driving major growth in the marketing industry. It is quickly becoming the top marketing assistant, with 64% of marketers already using it and 38% planning to start in 2004. Furthermore, 60% of respondents are concerned that AI could harm their brand's reputation through bias, plagiarism, or misalignment with brand values. AI tools are also aware of these problems and are already introducing possibilities for checking data and sources, for example, Google's Gemini (Hubspot, 2024).

2. RISKS OF INFLUENCER MARKETING

222 Influencer marketing has become an essential digital marketing strategy for marketing and profit generation. Nonetheless, influencer marketing isn't without risk. If clear instructions are not set for influencers, brands risk the campaign

losing its meaning. Additionally, careful selection of influencers is extremely important due to the risk of close and negative associations with individuals prone to personal indiscretions or scandals (Campbell et al., 2020). There are numerous examples of failed influencer marketing campaigns.

Perhaps one of the most famous is the example of the luxurious music Fyre Festival, which was conceived to be held in the Bahamas in 2017. The organizers hired hundreds of influencers, including celebrities like Kendall Jenner, Bella Hadid, and Emily Ratajkowski, to promote the event on social media creating a huge hype and attracting thousands of attendees. Attracted by the influencer's posts, attendees paid thousands of dollars for tickets and accommodation, promised white sand, luxury lodgings, and first-class food. However, visitors found themselves in the mud and under tents, without the security of the organization and with the canceled festival. This resulted in lawsuits and the main organizer being fined. Netflix made a documentary *FYRE: The Greatest Party That Never Happened* (2019). It turned out that the influencers who were raising the hype for the festival didn't even come to it. Thanks to the influencers, the tickets were sold, but they lacked transparency and ethics (Faster capital). Influencers also faced backlash from their followers and the public, who accused them of misleading them and being irresponsible.

Brands that opt for mega-influencers must also be aware of the risk of unpredictable reactions and personality traits of influencers. For example, shareholders of Tiger Woods's sponsors lost between \$5 and \$12 million following revelations of his infidelity and his arrest for driving under the influence of alcohol or drugs. Another risk is the huge focus on metrics, so influencers often pay bots to increase their number of followers or views.

Another potential danger is the discrimination of influencers. According to the Influencer Marketing Hub survey, 58.3% stated they had been discriminated against, most often regarding their gender, particularly on TikTok. Discrimination was felt by both macro and micro-influencers (Influencer Marketing Hub, 2023).

AI influencers behave similarly to real ones. They are considered more ethical than traditional influencers due to their involvement in socially significant

issues and socially responsible behavior. Caution should be exercised because of the possible disadvantages of blurring the lines between reality and fantasy, which can negatively affect consumers (Laszkiewicz, 2024).

A potential risk is also the appearance of an influential person. Virtual influencers have gender, skin color, language, and age, which affect their perception. In this context, brands should consider how much their brand image depends on the correct selection of virtual influencers that should influence people's perceptions (Influencer Marketing Hub, 2023:43).

If they have many followers, they are sure to attract the attention of companies with their popularity. However, questions arise regarding intellectual property, ethics, etc., and lack of authenticity, for example, how can influencers promote a product if they are unable to touch or use it? To ensure endorsements are not based solely on mimicking trends, both influencers and brands should find a way to deal with inauthentic endorsements to run a successful campaign (Kádeková and Holienčinová, 2018).

For companies to successfully master digital marketing based on artificial intelligence, they will need to consider legal and ethical considerations and ensure a safe and reliable environment for consumers (Yoo and Piscarac, 2023).

Another brand risk is the cancel culture. Cancel culture was born out of social media activism (Dotson and Ashloch, 2023) and can be described as a choice to withdraw one's attention from someone or something whose values, (in)action, or speech are so offensive, one no longer wishes to grace them with their presence, time, and money (Clark 2020, 85). Cancel culture has heavily influenced consumer-brand relationships in recent years (Pušić and Vojvodić, 2021:228), and the usual targets are brands and celebrities (Pušić and Vojvodić, 2021:232).

One example is the luxury fashion brand Balenciaga, which in 2022 released its new holiday campaign featuring children posing in their bedrooms next to the brand's products lined up like toys. After a few days of quiet public backlash, the campaign began to draw angry criticism over images of children with teddy bears inspired by the S&M brand, raising accusations and comments

that Balenciaga was sexualizing children. Creative director, Demna Gvasalia, linked Balenciaga to social media with a strong digital strategy, thus approaching Generation Z to the brand (Gárgoles and Ambás, 2023). The social media flooded with hashtags #cancelbalenciaga and #boycottbalenciaga with millions of views. Celebrities and influencers also called out Balenciaga for the ad. This example illustrates how social media impacted the brand and its misstep.

According to Statista (Image 4), with millions of internet users scouring social media platforms daily for entertainment, inspiration, and product recommendations, it is no surprise that brands are harnessing the power of social media's most recognizable faces for promotion. As of April 2024, there were 5.44 billion internet users worldwide, which amounted to 67.1% of the global population. Of this total, 5.07 billion, or 62.6% of the world's population, were social media users. These users inevitably look up to social media influencers to guide their decision-making. In addition to the influencer's credibility, creating content that is valuable to the audience is an important factor in gaining favor. Thus, the content should be useful in helping users make purchasing decisions. Moreover, the content that an influencer creates can be seen to help customers who are looking for information and can rely on someone they trust. Social media content is, therefore, a tool that can influence customer beliefs and purchasing behavior (Venciute et al. 2023). Given that the content produced by influencers is usually informative, entertaining, or inspirational, the expectation is that if followers find it useful, it will directly influence their purchase intent and decision.

Number of internet and social media users worldwide as of April 2024

(in billions)



Image 4: Number of internet and social media users worldwide as of April 2024 (in billions), available at <https://www.statista.com/statistics/617136/digital-population-worldwide/>

3. INFLUENCERS AND CONSUMER BEHAVIOR

Many studies confirm that GenZ's awareness about the impact of their purchasing decision is geared towards sustainability-oriented brands. As the youngest generation of active consumers on the market, GenZ is constantly online, analyzing, searching for options and ideas, participating and making their experience relevant for sharing with their virtual tribe. According to Kralj and Mamula (2022), companies need to attract the attention of GenZ members without relying on standard forms of communication and promotion by gathering them with their closest friends, which is easier to organize with the food delivery and shopping apps, to replace it with something innovative that requires leaving a comfort zone. One of the primary goals of a marketing campaign is to promote the brand and convince customers to use a product or service, leading to a

purchase decision (Mamula Nikolić et al., 2023). The content that influencers generate on social networks has a significant impact on consumer behavior.

Research confirms that 70% of teenage YouTube subscribers trust the opinions of influencers more than celebrities and 40% of millennials say their favorite YouTube personality understands them better than their friends (Venciute et al., 2023).

According to a survey report conducted in 2023 by the agency Matter Communications in the American states, as many as 81% of respondents answered that they bought, selected, or intend to buy products after the recommendation of influencers or family and friends who posted about that product (Deyo, Marketing Dive, 2023).

Influencers often share information from their private lives, and members of the new generation communicate with them more directly and more often. In a way, they see them as their online friends and role models (Mamula Nikolić et al., 2022).

By building the credibility and trust of followers, influencers gain significant influence over consumer purchasing behavior. By creating authentic and relatable content, using the latest digital technology, AI tools connect them with the audience which leads to making a purchase decision. If there is a lack of trust in influencers who do not instill confidence, the consumer will not buy the product or service.

Statista shows data from 2023, where the highest level of trust in brand and product recommendations from social media influencers was evident among Gen Z and millennials, with one-third of respondents from each generational cohort expressing complete trust. However, over 40% of baby boomers showed no confidence in influencer endorsements at all, nearly double the skepticism of Gen X and more than four times the skepticism of Gen Z (Statista, 2023). Research conducted by Sprout Social shows that 46% of Gen Z respondents expressed increased interest in brands collaborating with AI-generated influencers (Sprout Social, 2024).

Millennials helped shape the world of influencers by pioneering the concept of influencer marketing (Singer et al., 2023). GenY's proficiency in using almost

all digital devices has led to them buying mostly online from brands they trust, have read good reviews on the internet, and are well informed about products and services. They pay special attention to the impact that brands have on the environment. They also take care of personal data protection. According to the same Statista report, as of March 2023, nearly half of millennials worldwide found endorsements from influencers about a brand or product more attractive than regular advertisements. For 36% of respondents, influencer recommendations were also considered more reliable than average ads (Statista, 2023).

With the constant growth and expansion of influencer marketing comes a new trend in influencer marketing. The impact of artificial intelligence is reflected in the increasing use of artificial intelligence tools by brands. Brands use it mainly to choose the most suitable person for an influencer campaign and to track the campaign's success (Liu, 2021). Recently, companies have created AI-Generated Avatar Influencers with great success. Virtual influencers have clear benefits when it comes to online engagement and marketing. They don't age, they're free from (real) scandals and they can be programmed to speak any language. It's no surprise a number of companies and celebrities have caught onto the trend.

A new trend of de-influencing was created on TikTok in 2023 due to fake promotions by influencers to encourage brands and influencers to use more honest advertising. De-influencers offer often better and more honest alternatives to the reviewed products (Plazibat and Marunica, 2024).

CONCLUSION

In the paper, we investigated the role of influencers in the era of AI expansion and the digital revolution on consumer behavior. The findings reveal that in the era of digital transformation, there is an integration of digital technologies and artificial intelligence. Influencers play a pivotal role in shaping consumer behavior. Influencers are keeping pace with the expansion of AI, using AI tools, and have become essential in modern marketing strategies, especially on social

media platforms. Consumer reliance on influencer recommendations was also accelerated by the COVID-19 pandemic, which accelerated the trend toward online shopping. Artificial intelligence technologies improve the effectiveness of influencers by providing personalized content and data-driven insights, resulting in higher engagement and conversion rates. We also found that the degree of trust in influencers is very important. This research highlights the high level of trust in influencers among Gen Z and Millennials, who are more likely to be influenced by social media personalities, while Gen X remains more cautious.

A new chapter in influencer marketing is the emergence of virtual influencers. They offer brands greater control over messaging and can engage with audiences in innovative ways. Examining recent trends, research shows that virtual influencers are preferred over human influencers and are considered equally trustworthy, yet they also bring new challenges that need addressing. Brands must adapt and innovate in a rapidly changing environment as digital technologies and artificial intelligence continue to evolve, and their impact on consumer behavior and marketing strategies will undoubtedly grow. Marketing professionals should prioritize identifying suitable influencers and use them to identify target audiences whose communications would be relevant to drive brand engagement. Additionally, consideration should be given to identifying the most effective strategies, opportunities, challenges, platforms, and content that content marketers will soon be using.

REFERENCES

1. Campbell, Colin, and Farrel Justine Rapp. 2020. "More than meets the eye: The functional components underlying influencer marketing". *Business horizons*, 63(4), 469–479.
2. Campbell, Colin, Sands, Sean, Ferraro, Carla, Tsao, Hsiu-Yuan (Jody), and Mavrommatis, Alexis. 2020. "From data to action: How marketers can leverage AI." *Business Horizons*, 63(2), 227–243

3. Clark, Meredith D. 2020. “Drag them: A brief etymology of so-called cancel culture”. *Communication and the Public*, 5(3–4), 88–92. <https://journals.sagepub.com/doi/epub/10.1177/2057047320961562>
4. Conversation (2023). <https://theconversation.com/virtual-influencers-meet-the-ai-generated-figures-posing-as-your-new-online-friends-as-they-try-to-sell-you-stuff-212001>
5. Deveau, Richelle., Griffin, Sonia Joseph, and Reis, Steve. 2023. “AI-powered marketing and sales reach new heights with generative AI.” *McKinsey*. (2023, March) <https://www.mckinsey.com/capabilities/growth-marketing-and-sales/our-insights/ai-powered-marketing-and-sales-reach-new-heights-with-generative-ai> (accessed 11.06.2024)
6. Deyo, Jessica. (2023 February). *Marketing Dive*. <https://www.marketingdive.com/news/influencer-marketing-success-matter-study-2023/643310/> (accessed 30.06.2024)
7. Dotson, Sydney and Ashloch, Mary. 2023. “The Influencer Generations. Millennials and Generation Z are Ditching the Typical 9–5 to Post on Social Media.” *Rowman & Littlefield*. Chapter 6.
8. https://books.google.com.sa/books?hl=en&lr=&id=T-SmEAAQBAJ&oi=fnd&pg=PA109&dq=cancel+culture+influencer+marketing&ots=vOwJm9iaum&sig=BNDLM-1BZCfLSOrM9CuWdd9Rrwo&redir_esc=y#v=onepage&q=cancel%20culture%20influencer%20marketing&f=false (accessed 11.06.2024)
9. Faster Capital, “The Case Studies and Examples of Successful and Unsuccessful Influencer Marketing Campaigns” <https://fastercapital.com/topics/the-case-studies-and-examples-of-successful-and-unsuccessful-influencer-marketing-campaigns.html> (accessed 11.06.2024)
10. Marr, Bernard. 2023. “How Online Influencers and Idols Are Using Generative AI.” Available at *Forbes*. <https://www.forbes.com/sites/bernardmarr/2023/12/01/how-online-influencers-and-idols-are-using-generative-ai/> (accessed 10/06/2024)
11. Gárgoles, Paula, and Ambás, Gabriela. 2023. “The Power of Consumers on Social Media: A Case Study of Balenciaga’s Crisis Communication.”

- In International Conference on Fashion communication: between tradition and future digital developments* (pp. 3–13). Cham: Springer Nature Switzerland.
12. Gerlich, Michael. 2023. “The power of virtual influencers: Impact on consumer behaviour and attitudes in the age of AI.” *Administrative Sciences*, 13(8), 178.
 13. Govindan, Priya, and Alotaibi, Ibrahim. (2021, January). Impact of Influencers on Consumer Behaviour: empirical study. In *2021 2nd International Conference on Computation, Automation and Knowledge Management (ICCAKM)* (pp. 232–237). IEEE.
 14. HubSpot. 2023. “Marketing Trends in 2023 from AI to Z.” <https://www.hubspot.com/hubfs/2023%20State%20of%20Marketing%20Report.pdf> (accessed 11.06.2024)
 15. Hubspot. 2024. “The 2024 State of Marketing & Trends Report: Data from 1400+ Global Marketers.” <https://blog.hubspot.com/marketing/hubspot-blog-marketing-industry-trends-report> (accessed 29.06.2024)
 16. Influencermarketinghub. 2024. “The State of Influencer Marketing 2024: Benchmark Report.” <https://influencermarketinghub.com/influencer-marketing-benchmark-report/> (accessed 22.06.2024)
 17. Influencer Marketing Hub. 2023. “The State of AI in Influencer Marketing: A Comprehensive Benchmark Report.” 2023, p. 43
 18. Kádeková, Zdenka, and Holienčinová, Mária Holotová. 2018. “Influencer marketing as a modern phenomenon creating a new frontier of virtual opportunities.” *Communication Today*, 9(2).
 19. Kralj, Svetlana, and Mamula Nikolić, Tatjana. 2022. “Uloga nove generacije u budućim trendovima održivog poslovanja.” *XXIX naučni skup TEHNOLOGIJA, KULTURA I RAZVOJ*, Tivat, Montenegro, CIP – Каталогизација у публикацији ISBN 978-86-82183-21-1 (29; 2022; Тиват) COBISS.SR-ID 11224704
 20. Łaszkiewicz, Anna. 2024. “ARTIFICIAL INTELLIGENCE (AI) INFLUENCERS IN INFLUENCER MARKETING.” *Scientific Papers of Silesian University of Technology. Organization and*

- Management Series 2024, no. 193 (2024): 23–34. <http://dx.doi.org/10.29119/1641-3466.2024.193.2>.
21. Liu, Siti. (2021) “The impact of influencer marketing on brand engagement: A conceptual framework.” In *2021 4th International Conference on Humanities Education and Social Sciences (ICHESS 2021)* (pp. 2219–2224). Atlantis Press.
22. Mamula Nikolić, Tatjana, Popović-Pantić, Sanja, and Perčić, Karolina. 2023. “Technology brands and the role of GenZ in their co-creation.” In *Kucharska, W. & Lechman, E. Technology Brands in the Digital Economy*. Routledge, pp. 108–124. <https://doi.org/10.4324/9781003184638-8>
23. Mamula Nikolić, Tatjana, Perčić, Karolina, and Nećak, Mirjana. 2022. “Designing sustainable brand by engaging brand- consumers’ collaboration. Generation Z co-creators and influencers in Serbia.” *Marketing 1*(53) p. 37–50
24. Plazibat, Ivana, and Marunica, Stipe. 2024. “The impact of artificial intelligence and of the #deinfluencing trend on influencer marketing.” In *Economic and Social Development, Book of Abstracts, Dubrovnik, April*.
25. Pušić, Ana, and Vojvodić, Katja. 2021. “To be or not to be cancelled: tackling the challenges of cancel culture in brand management.” In *27th CROMAR Congress* (pp. 227–246).
26. Saini, Shweta, Ngah, Abdul Hafaz, Sahai, Seema, and Bansal, Rohit. (2024). “When AI Meets Influence: Exploring the Integration of ChatGPT and Influencer Marketing.” In *Leveraging ChatGPT and Artificial Intelligence for Effective Customer Engagement* (pp. 272–284). IGI Global.
27. Sands, Sean, Campbell, Collin, Plangger, Kirk, and Ferraro, Carla. 2022. “Unreal influence: Leveraging AI in influencer marketing.” *European Journal of Marketing*, 56(6), 1721–1747.
28. Singer, Mara, Callendar, Chaz, Ma, Xiao., & Tham, Samuel. 2023. “Differences in perceived influencer authenticity: a comparison of Gen Z and Millennials’ definitions of influencer authenticity during the

- de-influencer movement.” *Online Media and Global Communication*, 2(3), 351–378.
29. Statista. 2023. “Trust level in product and brand recommendations by social media influencers worldwide in 2023 by generational cohort.” Available at <https://www.statista.com/statistics/1455067/trust-in-influencers-for-brand-recommendations/> (accessed 30.06.2024)
30. Statista. 2023. <https://www.statista.com/statistics/1410704/millennials-trust-influencers-brand-recommendation-worldwide/> (accessed 11.06.2024)
31. Sproutsocial. 2024. “The 2024 Influencer Marketing Report: In Influencers We Trust.” Available at <https://sproutsocial.com/insights/data/influencer-marketing-report1/> (accessed 30.06.2024)
32. Venciute, Dominyca, Mackeviciene, Ieva, Kušlys, Marius, and Correia, Ricardo. (2023). “The role of influencer–follower congruence in the relationship between influencer marketing and purchase behaviour.” *Journal of Retailing and Consumer Services*, 75(1), 103506.
33. Yoo, Seung-Chul, and Piscarac, Diana. 2023. “Generative AI and its Implications for Modern Marketing: Analyzing Potential Challenges and Opportunities.” *The International Journal of Advanced Smart Convergence*, 12(3), 175–185

PROGRESS OF ARTIFICIAL INTELLIGENCE: CHALLENGES AND OPPORTUNITIES IN E-TOURISM

Žaklina Spalević¹

Singidunum University, Belgrade

Ivan Trifunović²

Tourist Organization of Vrnjačka Banja, Vrnjačka Banja

Sanja Marković³

*Academy of Vocational Studies of Kosovo and Metohija – Department
Zvečan*

Abstract: *In recent years, travel and tourism have undergone significant changes influenced by globalization, digitalization, sustainability, and the COVID-19 pandemic. These factors have shaped the industry in various ways, presenting both challenges and new opportunities. Globalization has facilitated the mobility of people, while digitalization has transformed how travel is planned and booked. Interest in travel continues to grow, especially due to the increasing availability of information online. All these factors together contribute to the significant growth of the travel and tourism market. The widespread use of computers, laptops, and smart devices has enabled a growing number of*

1 zspalevic@singidunum.ac.rs

2 ivantrifunovic75@gmail.com

3 sanja.markovic@akademijakm.edu.rs

people to use the Internet for various commercial purposes. Thanks to mobile applications that perform tasks previously requiring personal computers to complete via websites, people now possess greater technological knowledge. This technological advancement has eased access to various e-services. E-tourism, as a significant segment of electronic commerce, has seen a substantial increase in popularity worldwide. This paper explores how artificial intelligence impacts the tourism industry, offering new opportunities while simultaneously presenting challenges. We will show how this technological innovation is shaping tourism and examine the development of the artificial intelligence market, as well as the role of legal regulations in their application in tourism.

Keywords: *information and communication technologies, artificial intelligence, e-tourism, legal regulations.*

INTRODUCTION

The rapid development of information and communication technologies (ICT) has significantly altered the landscape of tourism in recent decades. The Internet, as a key platform for communication and information exchange worldwide, has transformed the way people travel. This evolution has created an interactive environment where travelers can customize their journeys according to their preferences. ICT and tourism are closely interconnected, and it is often difficult to distinguish where ICT drives or facilitates changes in the demand and supply of tourism services. While tourism generates increasing demands through the expansion of traveler needs and the growth of consumer demand, ICT provides tools that adapt to these needs, enabling development and increasing competition in the tourism sector (Ramos & Rodrigues, 2010, p. 1; Mavri & Angelis, 2009, p. 113; Frechtling, 2001, p. 154).

236 | The widespread availability and use of computers, laptops, and other smart devices have increased the number of people using the Internet for various

commercial purposes (De, Pandey & Pal, 2020, p. 1). In today's market, it is essential to talk about the Internet as a distribution channel. This makes direct sales from suppliers to consumers more feasible than ever before (Goeldner & Ritchie, 2009).

People now possess greater technological knowledge, as mobile applications perform tasks that were previously carried out using personal computers (PC) via websites (Deng, 2013, p. 351). Digital transformation, facilitated by the abundance of available data, has dramatically changed all aspects of consumer habits – from searching and booking travel to expectations during the trip itself. These changes have forced tourist organizations and companies to adapt to the new demands and opportunities of the digital age to meet the needs of modern travelers and remain competitive in the market (Aikaterini & Kamenidou, 2021). This advancement has made access to various e-services simpler. The use of technology has had a significant impact on all activities and the transformation of business practices, especially in the tourism industry.

The tourism industry is highly susceptible to changes brought about by technological advancements, primarily due to its pronounced focus on service provision. Therefore, tourist organizations and companies are compelled to adapt to these changes and leverage the opportunities provided by the digital age to meet all the demands and needs of contemporary travelers. This includes investing in digital technologies, educating employees on new tools and practices, and continuously monitoring and adjusting strategies in line with changes in consumer needs and preferences.

PROGRESS OF ARTIFICIAL INTELLIGENCE

Artificial Intelligence (AI) is a branch of computer science focused on developing algorithms and methods that enable machines to possess human-like intellectual capabilities, including learning, reasoning, and understanding (Russell & Norvig, 2010).

The advent of AI brings a series of challenges, including issues of privacy, data protection, and ethical dilemmas arising from the use of AI (Gerke, Minssen & Cohen, 2020, p. 295). With technological advancement and growing awareness of data privacy, managing personal information becomes increasingly complex. Therefore, it is crucial for companies in the hospitality and tourism sector to actively monitor and adjust their practices to ensure that users have confidence in how their data is processed and used (Kuner, 2018). Managing this data is challenging as it requires balancing the use of data to provide a better user experience and respect user privacy. To ensure that users trust how their data is used, companies must actively follow changes in data privacy legislation and implement appropriate data protection measures.

Thus, the European Union's (EU) approach to AI is based on the idea that individuals and companies can benefit from AI, but in a safe manner with respect to precise rules. This approach aims to ensure the smooth functioning of the market and the public sector, as well as the protection of individuals' fundamental rights. Through this approach, individuals and companies could develop trust in AI, which would further enable its advancement. The path to creating a legal framework for AI in the EU has not been quick or easy (Prlja, Gasmi & Korać, 2021, p. 57).

Legal systems must adapt to the challenges posed by the increasing use of AI. To ensure a high level of reliability and safety in the functioning of AI systems, it is necessary to adopt new precise regulations. These regulations should define standards that both high-risk and lower-risk AI applications must meet. Since many issues related to the use of AI have not been regulated so far, it is necessary to establish regulations that will allow companies and individuals to use AI safely and provide legal mechanisms for addressing any negative consequences of using this technology (Prlja, Gasmi & Korać, 2021, p. 97).

The EU AI Act is a comprehensive legislative instrument that applies to all AI systems introduced to the EU market, used within the EU, or used by EU users. A key part of this law is a categorization system that assesses the potential risks AI technology may pose to the well-being, safety, or fundamental rights

of individuals. This framework encompasses four risk levels: minimal, high, unacceptable, and specific risk of lack of transparency (EU in Serbia, 2023).

Minimal risk: Most AI systems fall into the minimal risk category. Applications with minimal risk, such as recommendation systems or spam filters, will not have strict obligations or restrictions as they do not pose significant risks to the rights or safety of users. Companies can voluntarily adopt additional codes of conduct for these AI systems.

High risk: AI systems identified as high-risk must comply with strict requirements, including risk mitigation, high-quality data sets, maintaining detailed documentation, providing clear information to users, human oversight, and a high level of robustness, accuracy, and safety. Regulatory frameworks will facilitate responsible innovation and the development of compliant AI systems. Examples of high-risk systems include critical infrastructures, medical devices, systems for access to educational institutions, and law enforcement.

Unacceptable risk: AI systems that clearly endanger fundamental human rights will be prohibited. This includes manipulative systems or applications, as well as the use of biometric systems that violate privacy or enable discrimination.

Specific risk of lack of transparency: Users must be aware when they are interacting with AI systems such as chatbots. Fake content and other AI-generated content must be clearly labeled. Providers will need to ensure transparency regarding synthetic content (European Commission, 2023).

“AI for Europe” is a strategic document adopted by the European Commission in April 2018, noting that AI has become a part of everyday life, from using virtual assistants to organize the workday to traveling in autonomous vehicles and using smartphones that suggest activities like listening to music and choosing restaurants for lunch. In this document, AI is defined as a system that demonstrates intelligent behavior by analyzing its environment and taking actions to achieve specific goals with a certain degree of autonomy. These systems can be purely software-based, operating in a virtual environment (e.g., image, speech, and face recognition systems), or embedded in hardware devices such as robots, drones, and autonomous vehicles. In many European countries,

AI is already used in various sectors. For example, in Denmark, it is used in emergency services to diagnose based on the sound of the patient's voice, while in Austria, it assists radiologists in more accurately detecting tumors. It is also used on farms across Europe to monitor the temperature and food consumption of animals. Overall, AI, as stated in this document, has become a key technology of the 21st century, with an increasing impact on various aspects of our lives and businesses (European Commission, 2023).

In today's digital age, AI represents a key component of technological advancement and societal transformation. This technology encompasses a wide range of algorithms and methodologies that enable machines to perform tasks traditionally requiring human intelligence, such as learning, reasoning, and decision-making. Through various sectors and applications, AI is becoming an indispensable part of our daily lives.

One of the key aspects of the AI market's development is the role of technological giants like Google, Microsoft, Amazon, and Apple. These companies invest substantial resources in AI research and development, contributing to the rapid advancement of this field. On the other hand, there is an increasing number of specialized firms providing AI solutions tailored to the needs of various industries, further stimulating market growth. Data supporting this fact is illustrated in Figure 1.

According to an analysis by Next Move Strategy Consulting (Next Move Strategy Consulting, 2024), the AI market is expected to see significant growth in the coming decade. Currently valued at nearly \$100 billion, it is projected to reach almost \$2 trillion by 2030. This impressive growth is driven by the increasing integration of AI into various industries, such as supply chains, marketing, manufacturing, research, analytics, and many other areas.

Key trends contributing to this growth include the development of chatbots, generative AI, and mobile applications. Chatbots are becoming increasingly popular as a means of automating customer support and communication. Generative AI, which enables the creation of new content such as texts, images, and sounds, is also being increasingly used in various applications, from creating artwork to personalized content production.

(in million U.S. dollars)

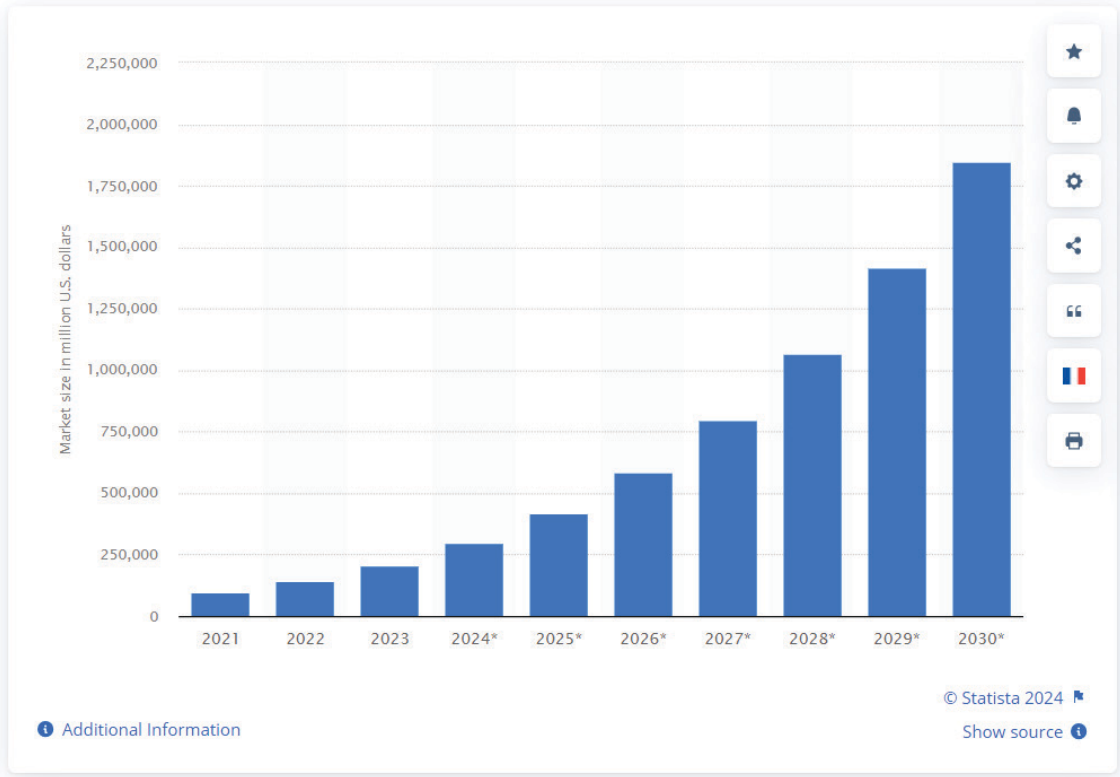


Figure 1: Global AI market size in 2021 with forecast to 2030 (Statista, 2023a)

The release of the ChatGPT 3.0 chatbot in 2022 marked a significant milestone for generative AI. This advanced platform allows for the creation of high-quality content on demand, opening up new possibilities in areas such as writing, design, and marketing. Interest in this field surged from 2022 to 2023, and this trend is expected to continue as ChatGPT and other tools are refined in the future, alongside the development of new generative AI programs.

With the continuous development of technology and the increasing implementation of AI in business processes, the AI market is expected to keep growing and expanding in the coming years. This will bring numerous opportunities and challenges for companies and organizations worldwide, as well as new ways to enhance efficiency and innovation in various sectors (Statista, 2023a).

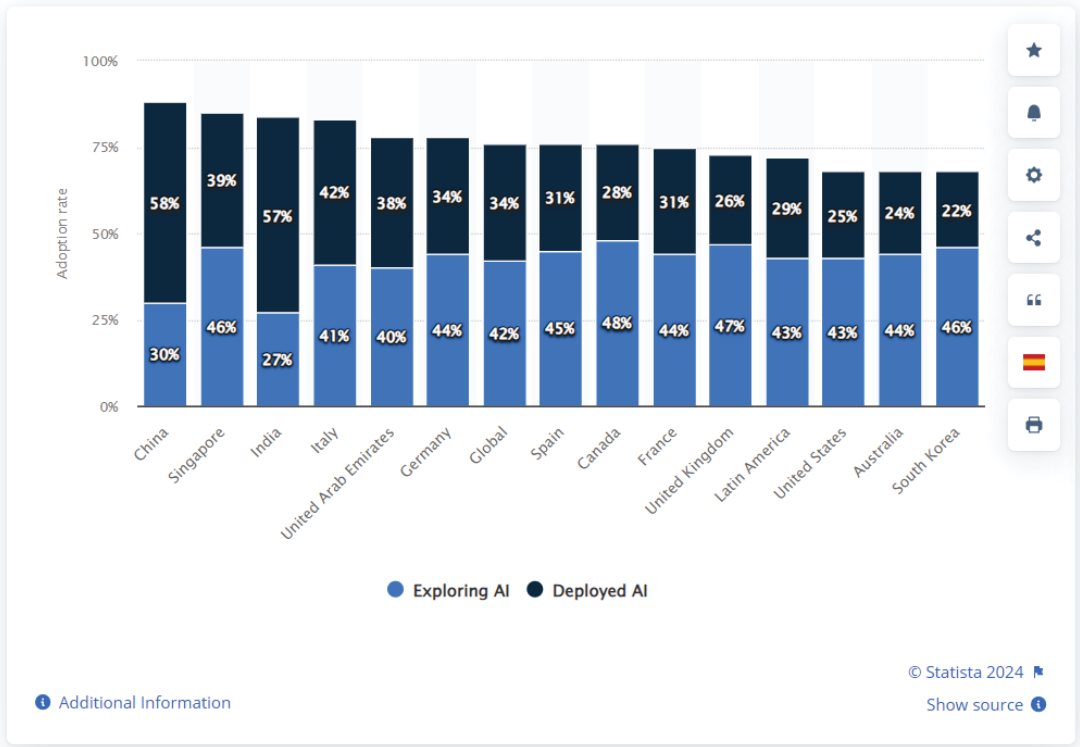


Figure 2: Global AI market size in 2021 with forecast to 2030 (Statista, 2023b)

According to the data in Figure 2, China was the global leader in AI research and application in 2022. This advantage was particularly evident in the implementation of AI technologies in organizations within China, where India ranked second. Both countries achieved an almost equal AI adoption rate of around 60%.

Focusing on research alone, the leading countries were Canada and the United States. These countries stood out for their high levels of activity in AI research and development, making them key players in the global AI landscape. Europe is also experiencing significant growth in the tech sector, especially in AI. Many leading European economies are at the forefront of AI research and application across various sectors. This highlights the growing importance of AI technologies in the global economy and shows that different regions of the world are increasingly committed to developing these technologies to enhance competitiveness and innovation.

In light of these trends and potential, the Republic of Serbia should continue to keep pace with the global development of AI by encouraging research, development, and application of this technology in various sectors. By the end of 2022, generative AI had become an important trend in AI development. Although this technology is not entirely new, as similar programs have been present over the past five years, the progress in this field has been remarkably impressive. Generative AI allows programs to create texts, images, music, and other content. The release of the ChatGPT-3 chatbot at the end of 2022 further spurred interest in this field, attracting the attention of companies such as Google, Microsoft, and Amazon. All these players are competing to develop the most advanced generative AI through various startup companies (Statista, 2024).

The company Oxford Insights, in collaboration with the International Development Research Centre (IDRC), has prepared the AI Readiness Index (Index), which is the only existing international indicator providing comparable measurements for a large number of countries worldwide. This composite Index means that the data forming the overall score is taken from other existing indices from previous years (some data dating back to 2016), as well as from publicly available data. Although the focus of the Index is on providing public services, the set indicators also reflect the maturity level of the private sector and society as a whole for AI-based technological solutions (Law on the Planning System of the Republic of Serbia, 2018).

According to the 2019 report, the Republic of Serbia was ranked 58th out of 194 countries, and in 2023 it was ranked 57th out of 193 countries. The Index

measures 11 indicators grouped into four areas: governance, infrastructure and data, skills and education, and public administration and services.

Table 1. Regional overview of AI readiness according to the Index

Country	Global ranking for 2019	Global ranking for 2023
Slovenia	38	39
Bulgaria	47	51
Hungary	48	45
Romania	55	64
Republic of Serbia	58	57
North Macedonia	61	83
Croatia	62	70
Montenegro	67	78
Albania	83	89
Bosnia and Herzegovina	95	117

From Table 1, we can see that from 2019 to 2023, the Republic of Serbia maintained a relatively stable position in the AI readiness ranking, with a slight improvement from 58th to 57th place. This data suggests a continuous effort by the state in this area but also highlights the challenges accompanying AI development in a global context. During the same period, other countries in the region experienced various changes in their positions. For instance, Slovenia, as the best-ranked country, also maintained a stable position with a slight decrease from 38th to 39th place. Hungary showed improvements, moving from 48th to 45th place, while North Macedonia, Croatia, Montenegro, and Bosnia and Herzegovina experienced significant drops in ranking. This trend reflects the dynamic nature of AI development in the region. Factors such as infrastructure, education, and governance play key roles in shaping a country’s readiness for AI integration into its society and economy. For example, a well-trained workforce and clear regulatory frameworks can facilitate the adoption and implementation

of AI. Additionally, investments in research and development, as well as cooperation between the public and private sectors, can foster innovation and enhance a country's competitiveness in this field.

The development of AI is an important factor in economic and social transformation. The Republic of Serbia, along with other countries in the region, faces the challenges and opportunities this process brings. Serbia's stable position in the AI readiness ranking indicates continuous efforts in this area but also reminds us of the need for further improvements in infrastructure, education, and regulatory frameworks to leverage all the advantages AI can bring.

The Republic of Serbia will invest 100 million euros in AI development over the next two years, supported by a new AI Development Strategy until 2030, along with an accompanying Action Plan. Serbia already adopted an AI Development Strategy in 2019, becoming the first country in Southeast Europe to have such a document. Serbia joined the Global Partnership on AI in 2022 and became a member of the World Economic Forum's AI Governance Alliance in 2023. According to the AI Readiness Index for 2023, Serbia ranks 57th out of 193 ranked countries, as a leader in the region (Naled, 2024).

The tourism sector is regulated by the Law on Tourism, the Law on Hospitality, and the Tourism Development Strategy for the period 2016–2025. Additionally, on July 23, 2023, the Government of Serbia adopted Ethical Guidelines for the Development, Application, and Use of Trustworthy and Responsible AI. These guidelines aim to ensure that science, particularly in the field of AI, develops and progresses without endangering or neglecting the human being, as the central figure of all processes that affect them and in which they are an indirect or direct participant. Furthermore, AI systems being developed must align with the well-being of humans, animals, and the environment.

To ensure the gradual alignment of the legislative framework with the legal acquis of the European Union, as well as the creation of a legal framework within the Republic of Serbia for the development and application of ethically aligned AI systems, the Government, through a Conclusion, adopts the document of Guidelines, recommending their application by all state bodies and organizations.

bodies and organizations of provincial autonomy, bodies and organizations of local self-government units, institutions, public enterprises, special bodies through which regulatory functions are performed, and legal and natural persons entrusted with public powers, when establishing and using AI systems in their work. It is also recommended that other legal and natural persons developing and/or using AI systems apply these Guidelines (Conclusion on the Adoption of Ethical Guidelines for the Development, Application, and Use of Trustworthy and Responsible Artificial Intelligence, 2023). Additionally, in the Republic of Serbia, the tourism sector is regulated by the Law on Tourism, the Law on Hospitality, and the Tourism Development Strategy for the period from 2020 to 2025.

ARTIFICIAL INTELLIGENCE IN ENHANCING TOURISM

The development of AI is having an increasingly significant impact on various sectors, including the tourism industry. Developed countries have highlighted tourism as a key priority to attract investments, increase revenues, create new employment opportunities, and encourage economic diversification (Alnafjan, Aldayel, and Kharrat, 2023, p. 1). An analysis of the available literature suggests that tourism has both direct and indirect impacts on a country's economy, primarily on its gross domestic product (GDP) and national income. The direct impact is reflected in the support of certain material production sectors that serve the tourism industry, while the indirect impact involves the transfer of income from other countries into the domestic economy (Trifunović, Spalević, Marković, and Savić, 2023).

By leveraging AI algorithms and machine learning techniques, organizations can achieve more accurate predictions and forecasts, enabling them to anticipate future trends, demand patterns, and market fluctuations. This facilitates informed decision-making and strategic planning, providing the ability to foresee future business conditions and revenues, as well as identify current and potential user/tourist demand trends (Doborjeh, Hemmington, Doborjeh, and Kasabov, 2022).

In the context of the tourism industry, machine learning algorithms can analyze vast amounts of data, including user preferences and behavior, to provide personalized recommendations for travel routes, accommodations, and activities. Additionally, AI can optimize operational efficiency by analyzing historical data on booking patterns, occupancy rates, and pricing trends. This enables revenue management and more efficient back-end operations, such as inventory management and supply chain logistics (Madurga and Mendez, 2023).

Therefore, AI technologies are transforming how travelers plan, book, and experience their trips, while simultaneously offering opportunities for the tourism industry to enhance its services and experiences. Analyzing the works of numerous authors addressing this issue, we will present several key aspects of the relationship between AI and the tourism industry, highlighting examples of AI technology application in practice.

Personalized travel experience: AI can be used to analyze traveler preference data to create personalized travel experiences. For instance, AI can recommend destinations, accommodations, activities, and restaurants based on the search history and past travels of the traveler.

Chatbot support: Tourism agencies and companies can use AI-powered chatbots to provide 24/7 support to travelers. Chatbots can answer questions about destinations, bookings, weather conditions, and other relevant information.

Travel data analysis: AI can analyze large volumes of travel data to identify trends, traveler preferences, and potential business improvement opportunities. For example, AI can identify popular destinations, travel seasons, or preferences for specific types of activities.

Predictive maintenance and resource management: In tourist destinations, AI can be used for predictive maintenance of infrastructure and resource management such as water, energy, and waste. This way, AI can help reduce costs and improve the efficiency and sustainability of destinations.

Marketing and personalized offers: Tourism agencies can employ AI to target specific groups of travelers and tailor marketing campaigns to their interests and preferences. For example, AI can analyze traveler data to create

personalized offers and promotions for specific market segments. AI enables tourism companies to customize marketing messages to users based on their preferences and past behavior. For example, AI algorithms can analyze user data to identify individuals likely interested in specific travel packages or destinations, and then send them personalized marketing messages (Giotis and Papadionysiou, 2022).

CONCLUSION

The fusion of ICT and AI in the tourism industry brings fundamental transformations in the process of planning, booking, and experiencing travel. This evolution allows personalized experiences for travelers, improves operational efficiency of enterprises, and opens doors for innovation. However, with technological advancement come challenges such as privacy issues, data protection, and ethical dilemmas arising from the use of AI. Therefore, it is essential to adopt precise regulations that ensure safe and responsible use of AI in tourism, while supporting innovation and protecting fundamental rights of individuals. Through a responsible approach and collaboration among industry, regulators, and academic communities, the full potential of technological advancement in the tourism industry can be realized, creating a sustainable and competitive global tourism scene.

Furthermore, the development of the AI market represents a global trend that is accelerating, with significant expected growth in the next decade. In this context, tech giants such as Google, Microsoft, Amazon, and Apple play a crucial role, investing substantial resources in the research and development of AI technologies. Additionally, an increasing number of specialized firms are providing AI solutions tailored to the needs of various industries, further stimulating market development. Key trends such as the development of chatbots, generative AI, and mobile applications are driving growth in this field. As the global AI market expands, countries like the Republic of Serbia are striving to keep pace with this trend and improve their readiness to integrate AI into society and the economy. The stable position of the Republic of Serbia in the readiness

ranking for AI indicates the efforts the country is making in this regard, while also emphasizing the need for further improvement in infrastructure, education, and regulatory framework to fully capitalize on the benefits that AI can bring.

REFERENCES

1. Aikaterini, S., & Kamenidou, I. (2021). „Complying with Digital Transformation in Online Booking through Experiential Values of Generation Z.“ *European Journal of Tourism Research*, 30(3003), 1–16.
2. Alnafjan, A., Aldayel, M., & Kharrat, A. (2023). „Systematic Review and Future Direction of Neuro-Tourism Research.“ *Brain Sciences*, 13(4), 682.
3. De, R., Pandey, N., & Pal, A. (2020). „Impact of digital surge during Covid-19 pandemic: A viewpoint on research and practice.“ *International Journal of Information Management*, 55, 1–5.
4. Deng, Z. (2013). „Understanding public users’ adoption of mobile health service.“ *International Journal of Mobile Communications*, 11(4), 351–373.
5. Doborjeh, Z., Hemmington, N., Doborjeh, M., & Kasabov, N.K. (2022). „Artificial Intelligence: A systematic review of methods and applications in hospitality and tourism.“ *International Journal of Contemporary Hospitality Management*, 34, 1154–1176.
6. EU u Srbiji. (2023). Godišnji izveštaj o Srbiji 2023. Retrieved from <https://europa.rs/godisnji-izvestaj-o-srbiji/>
7. European Commission. (2018). Artificial Intelligence for Europe. Retrieved from <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52018DC0237>
8. European Commission. (2023). Commission welcomes political agreement on Artificial Intelligence Act. Retrieved from https://ec.europa.eu/commission/presscorner/detail/en/ip_23_6473
9. Frechtling, D.C. (2001). *Forecasting Tourism Demand – Methods and Strategies*. Oxford: Linacre House.

10. Gerke, S., Minssen, T., & Cohen, G. (2020). „Ethical and Legal Challenges of Artificial Intelligence-Driven Healthcare.“ In *Artificial Intelligence in Healthcare*, 295–336.
11. Giotis, G., & Papadionysiou, E. (2022). „The Role of Managerial and Technological Innovations in the Tourism Industry: A Review of the Empirical Literature.“ *Sustainability*, 14(9), 5182.
12. Goeldner, R., & Ritchie, J. (2009). *Tourism: Principles, Practices, Philosophies*. New Jersey: John Wiley and Sons.
13. Kuner, C. (2018). „International Agreements, Data Protection, and EU Fundamental Rights on the International Stage: Opinion 1/ (EU-Canada PNR) of the Court of Justice of the EU.“ *Common Market Law Review*, 55(3), 857–882.
14. Madurga, M. A. G., & Mendez, A. J. G. (2023). „Artificial Intelligence in the Tourism Industry: An Overview of Reviews.“ *Administrative Sciences*, 13(8), 1–22.
15. Mavri, M., & Angelis, V. (2009). „Forecasting the Growth of e-Tourism Sector: The Case Study of Mediterranean countries.“ *Tourismos: An International Multidisciplinary Journal of Tourism*, 4(3), 113–125.
16. Naled. (2024). Srbija u razvoj veštačke inteligencije ulaže 100 miliona evra do 2026. Retrieved from <https://naled.rs/vest-srbija-u-razvoj-vestacke-inteligencije-ulaze-100-miliona-evra-do-2026-8731>
17. Next Move Strategy Consulting. (2024). Retrieved from <https://www.nextmsc.com/aboutus>
18. Prlja, D., Gasmi, G., & Korać, V. (2021). *Veštačka inteligencija u pravnom sistemu EU*. Monografija 177. Beograd: Institut za uporedno pravo.
19. Ramos, C., & Rodrigues, P. (2010). „The importance of online tourism demand.“ 10th International Forum on Tourism Statistics, Session 1: Statistical coverage of new trends in tourism, 1–20.
20. Russell, S.J., & Norvig, P. (2010). *Artificial Intelligence: A Modern Approach* (3rd ed.). Upper Saddle River: Prentice Hall.

21. Statista. (2023). Artificial intelligence (AI) market size worldwide in 2021 with a forecast until 2030. Retrieved from <https://www.statista.com/statistics/1365145/artificial-intelligence-market-size/>
22. Statista. (2023b). Adoption rate in business of AI worldwide and selected countries 2022. Retrieved from <https://www.statista.com/statistics/1378695/ai-adoption-rate-selected-countries/>
23. Statista. (2024). Artificial intelligence (AI) worldwide – statistics & facts. Retrieved from <https://www.statista.com/statistics/1378695/ai-adoption-rate-selected-countries/>
24. Trifunović, I., Spalević, Ž., Marković, F., & Savić, S. (2023). „Sinergija University Scientific Conference with International Participation.“ 35–40.
25. Zaključak o usvajanju etičkih smernica za razvoj, primenu i upotrebu pouzdane i odgovorne veštačke inteligencije („Sl. glasnik RS“, br. 23/2023). Nacionalna platforma za veštačku inteligenciju. Etičke smernice za razvoj, primenu i upotrebu pouzdane i odgovorne veštačke inteligencije. Retrieved from http://demo.paragraf.rs/demo/combined/Old/t/t2023_03/SG_023_2023_007.htm
26. Zakon o planskom sistemu Republike Srbije („Sl. glasnik RS“, br. 30/18). Strategija razvoja veštačke inteligencije u Republici Srbiji za period 2020–2025. godine. Retrieved from https://www.srbija.gov.rs/extfile/sr/437304/strategija_razvoja_vestacke_inteligencije261219_2_cyr.pdf

UDC: 004.8:338.48

343.85:343.53

DOI: 10.69899/limes-plus-en-24212-3253t

Original scientific work

APPLICATION OF ARTIFICIAL INTELLIGENCE IN DETECTING FRAUD IN TOURISM

Ivan Trifunović¹

Tourist Organization of Vrnjačka Banja, Vrnjačka Banja

Žaklina Spalević²

Faculty of Tourism and Hospitality Management, Belgrade

Dejan Rančić³

Faculty of Electronic Engineering, Niš

Filip Marković⁴

Faculty of Technical Sciences, Kosovska Mitrovica

Milan Simić⁵

RMIT University, Australia

Abstract: *Tourism has become a key sector of the economy in many countries during this century, offering immense potential for job creation and economic growth. The development of information technologies and artificial intelligence is transforming the tourism industry, enabling personalized experiences, travel optimization, and enhanced customer support. With the rise of online payments in tourism, the risk of various forms of fraud also increases.*

1 ivantrifunovic75@gmail.com

2 zspalevic@singidunum.ac.rs

3 dejan.rancic@elfak.ni.ac.rs

4 filip.markovic@pr.ac.rs

5 milan.simic@rmit.edu.au

Therefore, the role of artificial intelligence, particularly machine learning and deep learning is becoming increasingly important in detecting fraud and ensuring secure financial transactions in e-tourism. While artificial intelligence continues to improve the tourism industry, it also faces challenges related to online fraud, offering opportunities for enhanced customer support, personalized experiences, and improved efficiency while addressing security issues in online transactions.

Keywords: *tourism, artificial intelligence, online payments, machine learning, deep learning.*

INTRODUCTION

At the beginning of this century, tourism became one of the most dynamic sectors of economic activity worldwide, simultaneously representing a key job generator. Economically, tourism is an important factor in the recovery of national economies of countries with significant tourism resources that know how to properly use them. Considering the practically unlimited tourism resources, tourism is regarded as a sector with exceptional potential for long-term development. State tourism organizations play a crucial role as distributors of accurate, updated, and relevant information for potential tourists and the tourism industry as a whole. On the other hand, the state's role is to collect information and use it to formulate laws, policies, and strategies that support tourism development. The development of information technologies represents the future of the tourism industry, enabling fast distribution and efficient use of information. Owners of well-organized and trustworthy websites gain significant benefits. It is important to ensure the protection of users who access and purchase from such websites, which can be achieved through security and privacy measures provided by producers and distributors of tourism products via the internet, with state support to intervene in cases of non-compliance with national legislation (Nedelea & Bălan, 2010, p. 492).

The global tourism industry has experienced significant growth and attracted millions of travelers eager for new destinations and cultural experiences. However, a concerning and widespread phenomenon has emerged – fraud in tourism. Tourism fraud encompasses various deceptive and illegal activities targeting tourists, including fake services, scams, and identity theft. These activities pose significant threats to the financial well-being, personal safety, and overall experience of travelers.

Fraud in tourism takes various forms and exploits the vulnerability and trust of tourists who are often unfamiliar with the destinations they visit. These fraudulent activities range from misrepresentation of accommodations, tour packages, or attractions to fake online reservations and financial transactions. As a result, tourists face the consequences of fraud, which range from minor inconveniences to severe financial losses and concerns for personal safety (Corpuz, Manlutac, De Guzman, & Santos, 2023, p. 154).

Global perspectives on tourism fraud may vary, but common themes and issues are internationally recognized. Tourism fraud refers to various deceptive practices targeting tourists with the aim of unfairly obtaining money or personal data. This can include activities such as fake tours, identity theft, and cybercrime. With the rise of online bookings, the number of fraudulent websites and ads increases, leading to financial losses for tourists (Corpuz, Manlutac, De Guzman, & Santos, 2023, p. 154).

Artificial intelligence (AI) is significantly transforming the tourism industry by enabling personalized experiences, travel optimization, and enhanced customer support. AI algorithms assist in making decisions about destinations, transportation, accommodation, and activities, tailoring offers to the specific needs of tourists. At the same time, the development of digital technologies and the increase in online payments elevate the risk of financial fraud. The use of machine learning, deep learning, and other AI technologies is becoming crucial for recognizing and preventing fraud, thus protecting users and ensuring transaction security. This paper explores the role of AI in e-tourism and emphasizes the importance of AI in preventing fraud in online payments.

THE ROLE OF ARTIFICIAL INTELLIGENCE IN MODERNIZING THE TOURISM INDUSTRY

AI is particularly significant for travel and tourism for several reasons. Tourists face many decisions about their future travels, such as choosing destinations, transportation, accommodation, and activities. These decisions significantly impact their satisfaction during the trip. However, the vast number of available destinations, transportation means, accommodations, and activities creates an almost infinite number of options, requiring some form of assistance. Travel organizations and agents also face the challenge of finding travel packages tailored to the specific needs of users that best suit them. With a nearly unlimited number of potential clients, organizations must perfectly match supply and demand, a complex process suited to AI capabilities (Bulchand-Gidumal, Secin, O'Connor, & Buhalis, 2023, p. 1).

AI is increasingly impacting e-tourism, providing personalized experiences to travelers, faster responses to inquiries, and travel optimization. AI algorithms play a growing role in the development and enhancement of e-tourism. They enable process automation, the analysis of large volumes of data, and the provision of personalized services to tourists. However, while AI brings many advantages, its application also presents challenges, including ethical and security aspects that must be carefully considered (Spalević, Milosavljević, & Marković, 2024, p. 209).

AI is significantly transforming the tourism industry through various aspects (Prahadeeswaran, 2023, p. 12), such as:

Customer support: Chatbots and virtual assistants enable quick responses to inquiries, personalized recommendations, and booking assistance, improving the efficiency and responsiveness of customer support.

Marketing and personalization: AI analyzes large amounts of data to understand traveler preferences, enabling the creation of targeted marketing campaigns and dynamic pricing strategies to optimize revenue.

Demand forecasting: AI algorithms predict demand for tourist destinations, accommodation, and services, helping optimize resources, prices, and inventories.

Sustainable tourism: AI systems help manage energy consumption, and waste management, and promote eco-friendly transportation options, reducing the ecological footprint of tourism.

Smart destinations: AI technologies enable efficient traffic management, crowd control, and information dissemination to visitors, enhancing the visitor experience in real time.

Translation: AI-powered translation tools remove language barriers for international travelers, making tourism more accessible.

Cultural heritage preservation: Aplikacije AI applications like augmented reality (AR) and virtual reality (VR) help preserve and promote cultural landmarks through interactive content.

Crisis management: AI systems monitor weather conditions and natural disasters, helping ensure traveler safety and continuity of tourism operations.

Accessibility: AI improves accessibility for travelers with disabilities through voice-activated services and navigation apps.

Ethical considerations: As AI becomes more prevalent in tourism, it is important to address issues of data privacy, algorithmic bias, and job displacement to ensure fairness and sustainability.

AI continues to reshape the tourism industry, enhancing customer support, personalizing experiences, promoting sustainability, and improving efficiency.

TOURISM AND ONLINE PAYMENTS

Tourism and online payments have become closely linked with the development of digital technology and the growing popularity of travel. Online payments allow tourists to book accommodations, purchase airline tickets, pay for tours and activities, and various services and products during their travels, all via the Internet. This has brought numerous advantages, including greater convenience, speed, and ease in travel planning and booking processes, as well

as a wider choice of offers and destinations available to travelers worldwide. Travelers can explore different options, compare prices, and customize their trips to their needs, all from the comfort of their homes or with the help of a smartphone. However, the increase in online payments in tourism also comes with an increased risk of fraud. Fraudsters may attempt to exploit vulnerabilities in online payment systems to commit fraud, including identity theft, false advertising of tourist offers, or misrepresentation of services.

In recent years, the number of online payments has dramatically increased globally. With the growing popularity of e-commerce, mobile payment apps, and digital wallets, consumers are increasingly using the Internet for their financial transactions. However, along with this rise comes an increased risk of online fraud.

The digital payments market had a global transaction value of \$8.35 trillion in 2022 and, according to Statista – Market Insights, is the largest market within FinTech. The digitization of financial services is associated with disruptive changes in the industry concerning the billing process (online purchases) and the payment process at points of sale (offline purchases). Digital payments can therefore be considered the next evolutionary step that enables further financial services and replaces traditional, outdated payment methods (from cash and credit/debit cards to mobile and digital wallets). When it comes to digital payments, China is currently the largest market in the world, with a transaction value of \$3.639 trillion in 2022 (Statista, 2023).

According to the National Bank of Serbia, the purchase of goods and services via the Internet is becoming increasingly popular in the Republic of Serbia. In the first quarter of 2023, the number of dinar transactions via the Internet amounted to 8.7 million, representing an increase of 25.76% compared to the same period the previous year. The value of these transactions was 24 billion dinars, an increase of 38.11% compared to the first quarter of 2022. Additionally, in the same period, the number of transactions executed in euros increased by 29.96%, from 1.6 million to 2 million transactions, while their value rose by 36.49%, from 70.4 million euros to 96 million euros. Regarding transactions

executed in US dollars, in the first quarter of 2023, there was an increase in both the number of transactions, by a quarter (from 1.2 to 1.5 million), and their value, by 8% (from \$34.2 million to \$37 million), compared to the same period the previous year (National Bank of Serbia, 2023).

In 2022, e-commerce merchants worldwide faced various types of fraud, with social engineering tactics such as phishing, pharming, and whaling being the most common (Statista, 2024).

Phishing is a form of social engineering aimed at stealing an individual's personal information. It is usually carried out via fake emails that request the entry of sensitive information on a fraudulent website. These fake sites are designed to look authentic, posing a threat to users both at home and at work. Unlike phishing, which uses email for communication, pharming uses advanced techniques like stealthy DNS servers to manipulate traffic and redirect users to fraudulent websites. Criminals try to alter DNS servers to reroute users to fake locations. Pharming attacks often succeed on unprotected computers. Spear phishing is a form of phishing that targets specific individuals or organizations. Attackers tailor emails and SMS messages to target specific persons. Whaling is a particular form of spear phishing aimed at high-profile individuals such as executives (Issuu, 2021).

Thus, the aforementioned techniques target human errors, attempting to deceive individuals into sharing information or clicking on links that install malicious software on their devices. First-party misuse, such as friend fraud and refund fraud, ranked second, followed by card testing and identity theft. As if that weren't enough, online merchants have identified new fraud trends, including cases where fraudsters provide services that facilitate fraudulent activities for clients. Given that fraudsters are employing increasingly sophisticated strategies, it has become necessary for online merchants to effectively enhance their fraud prevention measures and counterattacks against these threats. A study showed that nearly three-quarters of companies planned to increase their budgets for fraud prevention in 2023. In 2022, tools such as Card Verification Number (CVN) checks and identity verification were the most commonly used for

fraud prevention. However, to strengthen their fraud management strategies and protect their businesses, online merchants are increasingly focusing on improving fraud analytics and the accuracy of automatic detection. Fraud affects the user experience. If merchants witness the financial impact of cybercrime, customers are not far behind them. In recent years, the percentage of online fraud victims who suffered financial losses has consistently remained above 70%. Since websites serve as the main channel for online fraud, this scenario inevitably damages merchants' reputations and undermines consumer trust. In 2022, seven out of ten global e-commerce users expressed a preference for payment methods that do not share their data with merchants, while nearly 60% expressed greater concern about online fraud during payment compared to the previous year. A study published a year earlier indicated that providing fraud protection could encourage the use of e-commerce. In the United States alone, about eight out of ten consumers would shop online more frequently if they were provided protection from these threats (Statista, 2024).

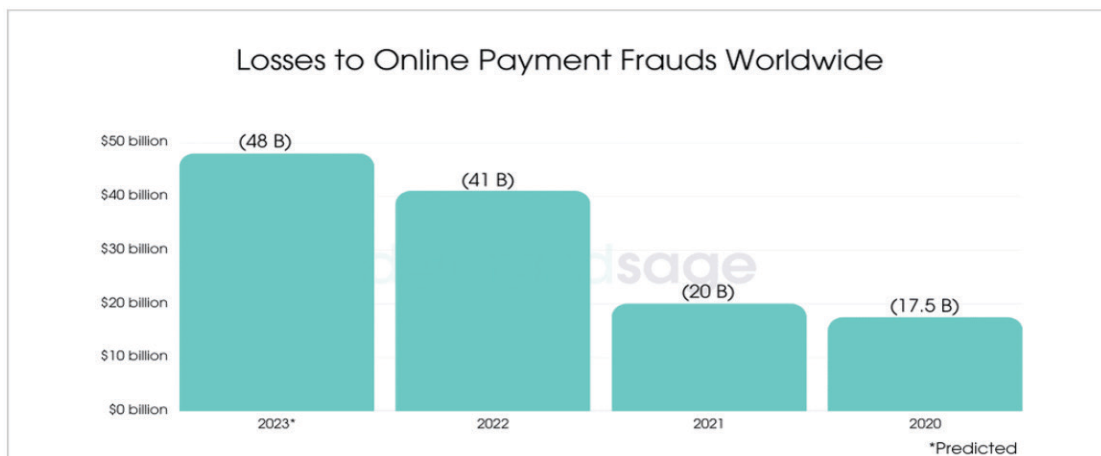


Figure 1. Losses due to online payment fraud worldwide (Ddemandsage, 2023)

260 Considering the growing popularity of online shopping and the increased number of transactions conducted electronically, the importance of AI in fraud detection is becoming ever more significant. With a greater number

of transactions comes a higher risk of fraud, and traditional fraud detection methods may be insufficiently effective or outdated. Thus, in the digital world, technology connects us and facilitates the exchange of information and services. However, with this progress comes new dangers. Fraud in financial transactions is becoming more frequent, causing significant losses and damage in the service sector. In the e-tourism sector, travelers often make online payments for accommodation, tickets, and other tourist services. Fraudulent activities are not limited to online transactions but can also occur in traditional offline payments. Fortunately, advanced technologies such as machine learning and AI, as well as classification algorithms, can help address these issues (Farzana, Onti, Islam, Islam, and Shatabda, 2023, p. 368).

ARTIFICIAL INTELLIGENCE IN FINANCIAL FRAUD PREVENTION

Many technologies we use in everyday life, from smart assistants (Google Assistant, Siri, etc.) to robotics and from facial recognition systems to autonomous vehicles, are examples of AI (Lu, Li, Chen, Kim, and Serikawa, 2018, p. 368). What makes these technologies AI are the components that comprise them. AI essentially encompasses multidisciplinary technologies, including machine learning, deep learning, computer vision, natural language processing, artificial neural networks, and expert systems (Hangl, Behrens, and Krause, 2022, p. 63).

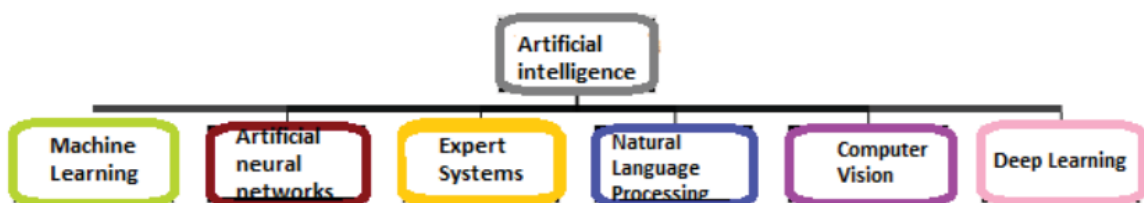


Figure 2. Sub-technologies of artificial intelligence (Hangl, Behrens, and Krause, 2022, p. 63)

Each of these sub-technologies has specific characteristics and applications that significantly contribute to the advancement of AI (Durmaz and Başer, 2023, p. 21).

Machine learning is the foundation of AI and encompasses algorithms that enable computers to learn and improve their performance based on experience. Machine learning is used in various industries for tasks such as fraud detection, credit scoring, recommendation systems, and medical diagnostics. Machine learning algorithms analyze large amounts of data, recognize patterns, and make decisions or predictions without being explicitly programmed for each task. This technology allows computers to become more autonomous and to adapt to new situations and information (Janiesch, Zschech, and Heinrich, 2021).

Deep learning, the most popular subfield of machine learning, uses multilayer neural networks to process complex information such as sound and images. Deep learning enables computers to learn from unlabeled data and extract high-level abstract information from it (Hao, 2019, p. 1).

Artificial neural networks are mathematical models inspired by the structure of the human brain and composed of interconnected artificial neurons that process information. The ambition to create a system that simulates the human brain spurred the initial development of neural networks. In 1943, McCulloch and Pitts attempted to understand how the brain could produce highly complex patterns using interconnected basic cells called neurons. The McCulloch-Pitts neuron model, known as the MCP model, laid the foundation for the later development of artificial neural networks (Voulodimos, Doulamis, Doulamis, and Protopapadakis, 2018, p. 1). These networks are capable of complex tasks such as pattern recognition, data classification, and prediction. Artificial neural networks are particularly useful in areas where traditional statistical techniques are not sufficiently precise, such as modeling financial time series and exchange rate forecasts (Abraham, 2005, p. 901).

262 | *Computer vision* is the science that studies how machines can “see” using computers and cameras. This technology enables computers to perceive, identify, measure, and track objects in the real world. Computer vision is used in

various applications, from facial recognition systems to automatic classification of microscopic images of cells. For example, the smart store “Amazon Go” uses computer vision technology based on deep learning to track customers and their purchases without the need for cashiers (Lu, 2019, p. 1).

Natural language processing (NLP) involves programming computers to properly process and understand human language. NLP facilitates communication between humans and machines, allowing computers to analyze, understand, and generate human language. Examples of this technology include chatbots that mimic human conversation and respond to customer inquiries. NLP has various applications, including text translation, sentiment analysis, and automatic text generation (Jain, Kulkarni, and Shah, 2018, p. 161).

Expert systems are intelligent computer systems that use expert knowledge to solve problems at a human level. These systems model the ability of humans to solve complex problems using predefined rules and knowledge bases. One of the first expert systems, “Dendral”, was developed at Stanford University for chemical analysis of Martian soil. Expert systems are used in medicine, engineering, finance, and many other fields where expert knowledge is needed for decision-making (Lucci and Kopec, 2016, p. 271).

Credit card fraud is an easy and attractive target. E-commerce and many other online sites have increased online payment methods, thereby increasing the risk of online fraud. With the rising fraud rates, researchers have begun to use various machine-learning methods to detect and analyze fraud in online transactions (Dornadula and Geetha, 2019, p. 631). Some of these methods include Decision Tree, Logistic Regression, Random Forest, Ada Boost, XGBoost, Support Vector Machine (SVM), and LightGBM (Lebichot, Paldino, Bontempi, Siblini, He-Guelton, and Oble, 2020, p. 785).

The application of AI and its aforementioned sub-technologies can significantly enhance the tourism industry’s ability to detect and prevent fraud, thereby increasing the security and trust of travelers. Here, we will emphasize machine learning and deep learning, which have been the most analyzed in the literature. Machine learning and deep learning algorithms can analyze

large volumes of transactions, reservations, and user behavior data to identify irregularities and anomalies that may indicate fraud. For example, they can detect frequent changes in travel reservations, unusually high expenses, or frequent purchase attempts from different locations, which may indicate suspicious activity.

Deep learning, including algorithms like autoencoders, plays a key role in improving security and trust in the tourism industry. Autoencoders are a special type of neural network used for anomaly detection, including credit card fraud. The main feature of autoencoders is encoding input data into a smaller number of features and then reconstructing them back into output data with the aim of accurately reconstructing the input (Misra, Thakur, Ghosh, and Saha, 2020, p. 254).

The autoencoder process consists of two basic parts (Pumsirirat and Yan, 2018, p. 18):

Encoder: The part of the network that compresses and reduces the dimensionality of the input data into a smaller number of features.

Decoder: The part of the network that reconstructs the original data from the compressed features.

When using the autoencoder model, the backpropagation algorithm is applied for error reconstruction. This algorithm adjusts the network weights to minimize the difference between the actual and reconstructed data.

The process of credit card fraud detection involves the following steps (Pumsirirat and Yan, 2018, p. 18):

Transaction collection: The issuing bank sends transactions to the user's bank with details such as amount, date, time, and location of credit card usage.

Behavior validation: Fraud detection systems use consumer profiles from the database to validate credit card behavior. Autoencoders train models using transaction history data and then use these models to validate new transactions.

Functionality of autoencoders: Autoencoders are designed to efficiently identify patterns and data structures without requiring pre-labeled transactions for training. This approach allows autoencoders to automatically detect

anomalies and irregularities in data, thereby contributing to increased security in the tourism industry.

The advantages of using autoencoders in fraud detection include (Pumsirirat and Yan, 2018, p. 18):

Unsupervised learning capability: Autoencoders do not require labeled data, enabling the use of large amounts of unlabeled data for training.

Efficiency in anomaly detection: Autoencoders are efficient in identifying patterns that deviate from normal behavior, which is crucial for identifying potential fraud.

Adaptability: They can be applied to different datasets and easily adapt to new types of fraud as they emerge.

In the tourism industry, autoencoders represent a key innovation promising to enhance the security and trust of travelers. Their ability to learn from unlabeled data enables the efficient use of a wide range of unlabeled information for training, while providing exceptional efficiency in detecting irregularities. This flexibility allows fraud detection systems to adapt and evolve to confront new challenges in the tourism world. Through such technological advancements, the industry can provide a safer travel experience and ensure that every traveler enjoys their journey without concerns about the security of their transactions.

CONCLUSION

Tourism is a sector that has experienced significant growth and development in recent decades, becoming a vital branch of many countries' economies. However, with this growth, new challenges have arisen, with one of the most significant being fraud in tourism. Fraudulent activities, especially in online payments, pose a serious threat to both financial stability and the reputation of the tourism industry.

In this context, the application of artificial intelligence has proven to be a crucial strategy for combating fraud and improving the security of online transactions. Through various techniques such as machine learning, deep learning,

and artificial neural networks, artificial intelligence can identify irregularities, recognize patterns, and automatically respond to potential fraudulent activities. Among all these techniques, deep learning stands out as one of the key technologies in combating fraud in the tourism industry, as it can analyze complex datasets and identify hidden patterns. This type of artificial intelligence enables systems to autonomously recognize the characteristics of fraudulent activities and adapt to new threats without the need for explicit programming.

The application of deep learning enables tourism companies to effectively detect suspicious activities and prevent financial losses, while also enhancing the traveler experience. This technology allows real-time monitoring of transactions and rapid response to potential threats, ensuring the security and reliability of tourism services for all users.

Artificial intelligence plays a crucial role in detecting fraud in tourism and improving the security of online payments, while also contributing to the modernization and efficiency of the tourism industry. With proper implementation and management, artificial intelligence can be a powerful force guiding tourism towards a more sustainable and secure future.

REFERENCES

1. Abraham, A. (2005). Handbook of Measuring System Design. 129 Artificial Neural Networks, Part 8. Elements: B – Signal Conditioning. John Wiley & Sons, Ltd. 901–908.
2. Bulchand-Gidumal, J., Secin, E. W., O'Connor, P., & Buhalis, D. (2023). Artificial intelligence's impact on hospitality and tourism marketing: exploring key themes and addressing challenges. *Current Issues in Tourism*, 1–18.
3. Corpuz, R. R. N., Manlutac, A. C., De Guzman, S. M. R., & Santos, C. J. P. (2023). Mitigating Tourism Fraud In Northern And Central Luzon: Understanding Factors Affecting Domestic Tourists And Enhancing

- Trust In Destination Experiences. *EPRA International Journal of Multidisciplinary Research (IJMR)*, 9(11), 154–168.
4. Ddemandsage. (2023). 59 eCommerce Fraud Statistics For 2024 (Latest Data). Retrieved from <https://www.demandsage.com/ecommerce-fraud-statistics/>
 5. Dornadula, V. N., & Geetha, S. (2019). Credit Card Fraud Detection using Machine Learning Algorithms. *Procedia Computer Science*, 165, 631–641.
 6. Durmaz, Y., & Başer, M. Y. (2023). A Systematic Literature Review on Artificial Intelligence Applications in Tourism Marketing. *International Journal of Research in Business Studies and Management*, 10(1), 21–30.
 7. Farzana, N., Onti, Y. L. H., Islam, T., Islam, M., & Shatabda, S. (2023). Fraud Detection Analysis of Revenue and Trip Transaction. 1–8.
 8. Hangl, J., Behrens, V. J., & Krause, S. (2022). Barriers, Drivers, and Social Considerations for AI Adoption in Supply Chain Management: A Tertiary Study. *Logistics*, 6(3), 63.
 9. Hao, Z. (2019). Deep learning review and discussion of its future development. *MATEC Web of Conferences*, 277, 02035: 1–7.
 10. Issuu. (2021). Phishing, pharming, whaling? Common data attacks and how to prevent them. Retrieved from https://issuu.com/castillians/docs/castille_newsletter_march_2021/s/12015870#google_vignette
 11. Jain, A., Kulkarni, G., & Shah, V. (2018). Natural Language Processing. *International Journal of Computer Sciences and Engineering*, 6(1), 161–167.
 12. Janiesch, C., Zschech, P., & Heinrich, K. (2021). Machine learning and deep learning. *Electronic Markets*.
 13. Lebichot, B., Paldino, G. M., Bontempì, G., Siblini, W., He-Guelton, L., & Oble, F. (2020). Incremental learning strategies for credit cards fraud detection: Extended abstract. *International Conference on Data Science and Advanced Analytics*, 785–786.

14. Lu, H., Li, Y., Chen, M., Kim, H., & Serikawa, S. (2018). Brain Intelligence: Go beyond artificial intelligence. *Mobile Networks and Applications*, 23(2), 368–375.
15. Lu, Y. (2019). Artificial intelligence: a survey on evolution, models, applications and future trends. *Journal of Management Analytics*, 6(1), 1–29.
16. Lucci, S., & Kopec, D. (2016). *Artificial Intelligence in the 21st Century, A Living Introduction*. (2nd ed.). Mercury Learning and Information, 271–272.
17. Misra, S., Thakur, S., Ghosh, M., & Saha, S. K. (2020). An Autoencoder Based Model for Detecting Fraudulent Credit Card Transaction. *Procedia Computer Science*, 167, 254–262.
18. Narodna banka Srbije. (2023). Saopštenje o podacima u vezi s pružanjem platnih usluga i izdavanjem elektronskog novca u prvom tromesečju 2023. godine. Retrieved from <https://www.nbs.rs/sr/scripts/showcontent/index.html?id=18941>
19. Nedelea, A., & Bălan, A. (2010). E-tourism and Tourism Services Consumer Protection. *Amfiteatru Economic*, 12(28), 492–503.
20. Prahadeeswaran, R. (2023). A Comprehensive Review: The Convergence of Artificial Intelligence and Tourism. *International Journal for Multidimensional Research Perspectives*, 1(2), 12–24.
21. Pumsirirat, A., & Yan, L. (2018). Credit Card Fraud Detection using Deep Learning based on Auto-Encoder and Restricted Boltzmann. *International Journal of Advanced Computer Science and Applications*, 9(1), 1–25.
22. Spalević, Ž., Milosavljević, B., & Marković, S. (2024). Legal Basis of Educational Processes of Artificial Intelligence Algorithms in E-tourism. *International Journal of Cognitive Research in Science, Engineering and Education*, 12(1), 209–217.
23. Statista. (2023). Top 100 companies: Online Payment. Preuzeto sa <https://www.statista.com/study/140117/top-100-companies-online-payment/>

24. Statista. (2024). E-commerce fraud - statistics & facts. Preuzeto sa [https://
www.statista.com/topics/9240/e-commerce-fraud/#topicOverview](https://www.statista.com/topics/9240/e-commerce-fraud/#topicOverview)
25. Voulodimos, A., Doulamis, N., Doulamis, A., & Protopapadakis, E. (2018). Deep Learning for Computer Vision: A Brief Review. Computational Intelligence and Neuroscience, 7068349, 1–13.